

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION  
International Bureau



BC

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : <b>C12Q 1/68</b>		<b>A2</b>	(11) International Publication Number: <b>WO 98/20166</b> (43) International Publication Date: <b>14 May 1998 (14.05.98)</b>																								
(21) International Application Number: <b>PCT/US97/20444</b> (22) International Filing Date: <b>6 November 1997 (06.11.97)</b> (30) Priority Data: <table><tr><td>08/744,481</td><td>6 November 1996 (06.11.96)</td><td>US</td></tr><tr><td>08/746,036</td><td>6 November 1996 (06.11.96)</td><td>US</td></tr><tr><td>08/746,055</td><td>6 November 1996 (06.11.96)</td><td>US</td></tr><tr><td>08/744,590</td><td>6 November 1996 (06.11.96)</td><td>US</td></tr><tr><td>08/786,988</td><td>23 January 1997 (23.01.97)</td><td>US</td></tr><tr><td>08/787,639</td><td>23 January 1997 (23.01.97)</td><td>US</td></tr><tr><td>08/933,792</td><td>19 September 1997 (19.09.97)</td><td>US</td></tr><tr><td>08/947,801</td><td>8 October 1997 (08.10.97)</td><td>US</td></tr></table> (71) Applicant (for all designated States except US): <b>SEQUENOM, INC. [US/US]; 11555 Sorrento Valley Road, San Diego, CA 92121 (US).</b> (72) Inventors; and (75) Inventors/Applicants (for US only): <b>KOSTER, Hubert [DE/US]; 8636 C Via Mallorca Drive, La Jolla, CA 92037 (US). TANG, Kai [CN/US]; 8521 Summerdale Road #241, San Diego, CA 92126 (US). FU, Dong-Jing [CN/US]; 10615 Dabney Drive #21, San Diego, CA 92126 (US). SIEGERT, Carston, W. [DE/US]; Geiselstrasse 42, D-22303 Hamburg (DE). LITTLE, Daniel, P. [US/US]; 393</b>			08/744,481	6 November 1996 (06.11.96)	US	08/746,036	6 November 1996 (06.11.96)	US	08/746,055	6 November 1996 (06.11.96)	US	08/744,590	6 November 1996 (06.11.96)	US	08/786,988	23 January 1997 (23.01.97)	US	08/787,639	23 January 1997 (23.01.97)	US	08/933,792	19 September 1997 (19.09.97)	US	08/947,801	8 October 1997 (08.10.97)	US	<b>Glendale Lake Road, Patton, PA 18668 (US). HIGGINS, G., Scott [GB/DE]; Haselweg 1, D-22880 Weidel (DE). BRAUN, Andreas [DE/US]; 13232 Benchley Road, San Diego, CA 92130 (US). DAMHOFFER-DEMAR, Brigitte [AT/US]; 3899 Haines Street #8-308, San Diego, CA 92109 (US). JURINKE, Christian [DE/DE]; Grope Hall 68, D-22115 Hamburg (DE). VAN DEN BOOM, Dirk [DE/DE]; Forsthausstrasse 8, D-63303 Prelech (DE). XIANG, Guobing [CN/US]; Apartment 23, 11381 Zapata Avenue, San Diego, CA 92126 (US). LOUGH, David, M. [GB/GB]; 32 Deanhead Road, Eyemouth, Berwickshire TD14 55A (GB).</b> (74) Agent: <b>SEIDMAN, Stephanie, L.; Brown Martin Haller &amp; McClain, 1660 Union Street, San Diego, CA 92101-2926 (US).</b> (81) Designated States: <b>AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).</b>  <b>Published</b> <i>Without international search report and to be republished upon receipt of that report.</i>
08/744,481	6 November 1996 (06.11.96)	US																									
08/746,036	6 November 1996 (06.11.96)	US																									
08/746,055	6 November 1996 (06.11.96)	US																									
08/744,590	6 November 1996 (06.11.96)	US																									
08/786,988	23 January 1997 (23.01.97)	US																									
08/787,639	23 January 1997 (23.01.97)	US																									
08/933,792	19 September 1997 (19.09.97)	US																									
08/947,801	8 October 1997 (08.10.97)	US																									
(54) Title: <b>DNA DIAGNOSTICS BASED ON MASS SPECTROMETRY</b> (57) Abstract <p>Fast and highly accurate mass spectrometry-based processes for detecting a particular nucleic acid sequence in a biological sample are provided. Depending on the sequence to be detected, the processes can be used, for example, to diagnose a genetic disease or chromosomal abnormality; a predisposition to a disease or condition, infection by a pathogenic organism, or for determining identity or heredity.</p>																											

# FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TC	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakhstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		
EE	Estonia						

-1-

## DNA DIAGNOSTICS BASED ON MASS SPECTROMETRY

### Related Applications

For U.S. National Stage purposes, this application is a  
5 continuation-in-part of U.S. application Serial No. 08/744,481, filed  
November 6, 1996, to Köster, entitled "DNA DIAGNOSTICS BASED ON  
MASS SPECTROMETRY". This application is also a continuation-in-part  
of U.S. application Serial Nos. 08/744,590, 08/746,036, 08/746,055,  
08/786,988, 08/787,639, 08/933,792 and U.S. application Serial No.  
10 atty dkt. no. 7352-2001B, filed October 8, 1997, which is a  
continuation-in-part of U.S. application Nos. 08/746,055, 08/786,988  
and 08/787,639. For international purposes, benefit of priority is  
claimed to each of these applications.

This application is related to U.S. Patent Application Serial No.  
15 08/617,256 filed on March 18, 1996, which is a continuation-in-part of  
U.S. application Serial No. 08/406,193, filed March 17, 1995, now U.S.  
Patent No. 5,605,798, and is also related U.S. Patent Nos. 5,547,835  
and 5,622,824.

Where permitted the subject matter of each of the above-noted  
20 patent applications and the patent is herein incorporated in its entirety.

### BACKGROUND OF THE INVENTION

#### Detection of mutations

The genetic information of all living organisms (e.g., animals,  
plants and microorganisms) is encoded in deoxyribonucleic acid (DNA).  
25 In humans, the complete genome is contains of about 100,000 genes  
located on 24 chromosomes (The Human Genome, T. Strachan, BIOS  
Scientific Publishers, 1992). Each gene codes for a specific protein,  
which after its expression via transcription and translation, fulfills a  
specific biochemical function within a living cell. Changes in a DNA  
30 sequence are known as mutations and can result in proteins with altered

-2-

or in some cases even lost biochemical activities; this in turn can cause genetic disease. Mutations include nucleotide deletions, insertions or alterations (i.e. point mutations). Point mutations can be either "missense", resulting in a change in the amino acid sequence of a protein  
5 or "nonsense" coding for a stop codon and thereby leading to a truncated protein.

More than 3000 genetic diseases are currently known (Human Genome Mutations, D. N. Cooper and M. Krawczak, BIOS Publishers, 1993), including hemophilias, thalassemias, Duchenne Muscular  
10 Dystrophy (DMD), Huntington's Disease (HD), Alzheimer's Disease and Cystic Fibrosis (CF). In addition to mutated genes, which result in genetic disease, certain birth defects are the result of chromosomal abnormalities such as Trisomy 21 (Down's Syndrome), Trisomy 13 (Patau Syndrome), Trisomy 18 (Edward's Syndrome), Monosomy X  
15 (Turner's Syndrome) and other sex chromosome aneuploidies such as Klienfelter's Syndrome (XXY). Further, there is growing evidence that certain DNA sequences may predispose an individual to any of a number of diseases such as diabetes, arteriosclerosis, obesity, various autoimmune diseases and cancer (e.g., colorectal, breast, ovarian, lung).  
20 Viruses, bacteria, fungi and other infectious organisms contain distinct nucleic acid sequences, which are different from the sequences contained in the host cell. Therefore, infectious organisms can also be detected and identified based on their specific DNA sequences.

Since the sequence of about 16 nucleotides is specific on  
25 statistical grounds even for the size of the human genome, relatively short nucleic acid sequences can be used to detect normal and defective genes in higher organisms and to detect infectious microorganisms (e.g., bacteria, fungi, protists and yeast) and viruses. DNA sequences can even serve as a fingerprint for detection of different individuals within th



same species (see, Thompson, J. S. and M. W. Thompson, eds., Genetics in Medicine, W.B. Saunders Co., Philadelphia, PA (1991)).

Several methods for detecting DNA are currently being used. For example, nucleic acid sequences can be identified by comparing the mobility of an amplified nucleic acid fragment with a known standard by gel electrophoresis, or by hybridization with a probe, which is complementary to the sequence to be identified. Identification, however, can only be accomplished if the nucleic acid fragment is labeled with a sensitive reporter function (e.g., radioactive ( $^{32}\text{P}$ ,  $^{35}\text{S}$ ), fluorescent or chemiluminescent). Radioactive labels can be hazardous and the signals they produce decay over time. Non-isotopic labels (e.g., fluorescent) suffer from a lack of sensitivity and fading of the signal when high intensity lasers are being used. Additionally, performing labeling, electrophoresis and subsequent detection are laborious, time-consuming and error-prone procedures. Electrophoresis is particularly error-prone, since the size or the molecular weight of the nucleic acid cannot be directly correlated to the mobility in the gel matrix. It is known that sequence specific effects, secondary structure and interactions with the gel matrix are causing artefacts.

#### 20 Use of mass spectrometry for detection and identification of nucleic acids

Mass spectrometry provides a means of "weighing" individual molecules by ionizing the molecules in vacuo and making them "fly" by volatilization. Under the influence of combinations of electric and magnetic fields, the ions follow trajectories depending on their individual mass (m) and charge (z). In the range of molecules with low molecular weight, mass spectrometry has long been part of the routine physical-organic repertoire for analysis and characterization of organic molecules by the determination of the mass of the parent molecular ion. In addition, by arranging collisions of this parent molecular ion with other

particles (e.g., argon atoms), the molecular ion is fragmented forming secondary ions by the so-called collision induced dissociation (CID). The fragmentation pattern/pathway very often allows the derivation of detailed structural information. Many applications of mass spectrometric methods are known in the art, particularly in biosciences (see, e.g., Methods in Enzymol., Vol. 193: "Mass Spectrometry" (J. A. McCloskey, editor), 1990, Academic Press, New York).

Because of the apparent analytical advantages of mass spectrometry in providing high detection sensitivity, accuracy of mass measurements, detailed structural information by CID in conjunction with an MS/MS configuration and speed, as well as on-line data transfer to a computer, there has been interest in the use of mass spectrometry for the structural analysis of nucleic acids. Recent reviews summarizing this field include K.H. Schram, "Mass Spectrometry of Nucleic Acid Components, Biomedical Applications of Mass Spectrometry" 34, 203-287 (1990); and P.F. Crain, "Mass Spectrometric Techniques in Nucleic Acid Research," Mass Spectrometry Reviews 9, 505-554 (1990); see, also U.S. Patent No. 5,547,835 and U.S. Patent No. 5,622,824).

Nucleic acids, however, are very polar biopolymers that are very difficult to volatilize. Consequently, mass spectrometric detection has been limited to low molecular weight synthetic oligonucleotides for confirming an already known oligonucleotide sequence by determining the mass of the parent molecular ion, or alternatively, confirming a known sequence through the generation of secondary ions (fragment ions) via CID in an MS/MS configuration using, in particular, for the ionization and volatilization, the method of fast atomic bombardment (FAB mass spectrometry) or plasma desorption (PD mass spectrometry). As an example, the application of FAB to the analysis of protected dimeric blocks for chemical synthesis of oligodeoxynucleotides has been

described (Köster et al. (1987) Biomed. Environ. Mass Spectrometry 14, 111-116).

Other ionization/desorption techniques include electrospray/ion-spray (ES) and matrix-assisted laser desorption/ionization (MALDI). ES mass spectrometry has been introduced by Fenn et al. (J. Phys. Chem. 88:4451-59 (1984); PCT Application No. WO 90/14148) and current applications are summarized in review articles (see, e.g., Smith et al. (1990) Anal. Chem. 62:882-89 and Ardrey (1992) Electrospray Mass Spectrometry, Spectroscopy Europe 4:10-18). The molecular weights of a tetradecanucleotide (see, Covey et al. (1988) The "Determination of Protein, Oligonucleotide and Peptide Molecular Weights by Ionspray Mass Spectrometry," Rapid Commun. in Mass Spectrometry 2:249-256), and of a 21-mer (Methods in Enzymol., 193, "Mass Spectrometry" (McCloskey, editor), p. 425, 1990, Academic Press, New York) have been published. As a mass analyzer, a quadrupole is most frequently used. Because of the presence of multiple ion peaks that all could be used for the mass calculation, the determination of molecular weights in femtomole amounts of sample is very accurate.

MALDI mass spectrometry, in contrast, can be attractive when a time-of-flight (TOF) configuration (see, Hillenkamp et al. (1990) pp 49-60 in "Matrix Assisted UV-Laser Desorption/Ionization: A New Approach to Mass Spectrometry of Large Biomolecules," Biological Mass Spectrometry, Burlingame and McCloskey, editors, Elsevier Science Publishers, Amsterdam) is used as a mass analyzer. Since, in most cases, no multiple molecular ion peaks are produced with this technique, the mass spectra, in principle, look simpler compared to ES mass spectrometry.

Although DNA molecules up to a molecular weight of 410,000 daltons have been desorbed and volatilized (Williams *et al.*, "Volatilization of High Molecular Weight DNA by Pulsed Laser Ablation of Frozen Aqueous Solutions," Science 246, 1585-87 (1989)), this technique had

5 only shown very low resolution (oligothymidylic acids up to 18 nucleotides, Huth-Fehre *et al.* Rapid Commun. in Mass Spectrom., 6, 209-13 (1992); DNA fragments up to 500 nucleotides in length K. Tang *et al.*, Rapid Commun. in Mass Spectrom., 8, 727-730 (1994); and a double-stranded DNA of 28 base pairs (Williams *et al.*, "Time-of-Flight

10 Mass Spectrometry of Nucleic Acids by Laser Ablation and Ionization from a Frozen Aqueous Matrix," Rapid Commun. in Mass Spectrom., 4, 348-351 (1990)). Japanese Patent No. 59-131909 describes an instrument, which detects nucleic acid fragments separated either by electrophoresis, liquid chromatography or high speed gel filtration. Mass

15 spectrometric detection is achieved by incorporating into the nucleic acids, atoms, such as S, Br, I or Ag, Au, Pt, Os, Hg, that normally do not occur in DNA.

Co-owned U.S. Patent No. 5,622,824 describes methods for DNA sequencing based on mass spectrometric detection. To achieve this, the

20 DNA is by means of protection, specificity of enzymatic activity, or immobilization, unilaterally degraded in a stepwise manner via exonuclease digestion and the nucleotides or derivatives detected by mass spectrometry. Prior to the enzymatic degradation, sets of ordered deletions that span a cloned DNA fragment can be created. In this

25 manner, mass-modified nucleotides can be incorporated using a combination of exonuclease and DNA/RNA polymerase. This permits either multiplex mass spectrometric detection, or modulation of the activity of the exonuclease so as to synchronize the degradative process.

-7-

Co-owned U.S. Patent Nos. 5,605,798 and 5,547,835 provide methods for detecting a particular nucleic acid sequence in a biological sample. Depending on the sequence to be detected, the processes can be used, for example, in methods of diagnosis. These methods, while  
5 broadly useful and applicable to numerous embodiments, represent the first disclosure of such applications and can be improved upon.

Therefore, it is an object herein to provide improved methods for sequencing and detecting DNA molecules in biological samples. It is also an object herein to provide improved methods for diagnosis of genetic  
10 diseases, predispositions to certain diseases, cancers, and infections.

#### SUMMARY OF THE INVENTION

Methods of diagnosis by detecting and/or determining sequences of nucleic acids that are based on mass spectrometry are provided herein. Methods are provided for detecting double-stranded DNA, detecting  
15 mutations and other diagnostic markers using MS analysis. In particular, methods for diagnosing neuroblastoma, detecting heredity relationships, HLA compatibility, genetic fingerprinting, detecting telomerase activity for cancer diagnosis are provided.

In certain embodiments the DNA is immobilized on a solid support  
20 either directly or via a linker and/or bead. Three permutations of the methods for DNA detection in which immobilized DNA is used are exemplified. These include: (1) immobilization of a template; hybridization of the primer; extension of the primer, or extension of the primer (single ddNTP) for sequencing or diagnostics or extension of the  
25 primer and Endonuclease degradation (sequencing); (2) immobilization of a primer; hybridization of a single stranded template; and extension of the primer, or extension of the primer (single ddNTP) for sequencing or diagnostics or extension of the primer and Endonuclease degradation (sequencing); (3) immobilization of the primer; hybridization of a double

stranded template; extension of the primer, or extension of the primer (single ddNTP) for sequencing or diagnostics or extension of the primer and Endonuclease degradation (sequencing).

In certain embodiments the DNA is immobilized on the support via  
5 a selectively cleavable linker. Selectively cleavable linkers include, but are not limited to photocleavable linkers, chemically cleavable linkers and an enzymatically (such as a restriction site (nucleic acid linker), a protease site) cleavable linkers. Inclusion of a selectively cleavable linker expands the capabilities of the MALDI-TOF MS analysis because it allows  
10 for all of the permutations of immobilization of DNA for MALDI-TOF MS, the DNA linkage to the support through the 3'- or 5'-end of a nucleic acid; allows the amplified DNA or the target primer to be extended by DNA synthesis; and further allows for the mass of the extended product (or degraded product via exonuclease degradation) to be of a size that is  
15 appropriate for MALDI-TOF MS analysis (i.e., the isolated or synthesized DNA can be large and a small primer or a large primer sequence can be used and a small restriction fragment of a gene or single strand thereof hybridized thereto).

In a preferred embodiment, the selectively cleavable linker is a  
20 chemical or photocleavable linker that is cleaved during the ionizing step of mass spectrometry. Exemplary linkers include linkers containing, a disulfide group, a leuvinyl group, an acid-labile trityl group and a hydrophobic trityl group. In other embodiments, the enzymatically  
25 cleavable linker can be a nucleic acid that is an RNA nucleotide or that encodes a restriction endonuclease site. Other enzymatically cleavable linkers include linkers that contain a pyrophosphate group, an arginine-arginine group and a lysine-lysine group. Other linkers are exemplified herein.

Methods for sequencing long fragments of DNA are provided. To perform such sequencing, specific base terminated fragments are generated from a target nucleic acid. The analysis of fragments rather than the full length nucleic acid shifts the mass of the ions to be

5 determined into a lower mass range, which is generally more amenable to mass spectrometric detection. For example, the shift to smaller masses increases mass resolution, mass accuracy and, in particular, the sensitivity for detection. Hybridization events and the actual molecular weights of the fragments as determined by mass spectrometry provide

10 sequence information (e.g., the presence and/or identity of a mutation). In a preferred embodiment, the fragments are captured on a solid support prior to hybridization and/or mass spectrometry detection. In another preferred embodiment, the fragments generated are ordered to provide the sequence of the larger nucleic acid.

15 One preferred method for generating base specifically terminated fragments from a nucleic acid is effected by contacting an appropriate amount of a target nucleic acid with an appropriate amount of a specific endonuclease, thereby resulting in partial or complete digestion of the target nucleic acid. Endonucleases will typically degrade a sequence into

20 pieces of no more than about 50-70 nucleotides, even if the reaction is not run to full completion. In a preferred embodiment, the nucleic acid is a ribonucleic acid and the endonuclease is a ribonuclease (RNase) selected from among: the G-specific RNase T<sub>1</sub>, the A-specific RNase U<sub>2</sub>, the A/U specific RNase PhyM, U/C specific RNase A, C specific chicken

25 liver RNase (RNase CL3) or crisavitin. In another preferred embodiment, the endonuclease is a restriction enzyme that cleaves at least one site contained within the target nucleic acid. Another preferred method for generating base specifically terminated fragments includes performing a combined amplification and base-specific termination reaction (e.g., using

-10-

an appropriate amount of a first DNA polymerase, which has a relatively low affinity towards the chain-terminating nucleotides resulting in an exponential amplification of the target; and a polymerase with a relatively high affinity for the chain terminating nucleotide resulting in base-specific termination of the polymerization. Inclusion of a tag at the 5' and/or 3' end of a target nucleic acid can facilitate the ordering of fragments.

Methods for determining the sequence of an unknown nucleic acid in which the 5' and/or 3' end of the target nucleic acid can include a tag are provided. Inclusion of a non-natural tag on the 3' end is also useful for ruling out or compensating for the influence of 3' heterogeneity, premature termination and nonspecific elongation. In a preferred embodiment, the tag is an affinity tag (e.g., biotin or a nucleic acid that hybridizes to a capture nucleic acid). Most preferably the affinity tag facilitates binding of the nucleic acid to a solid support. In another preferred embodiment, the tag is a mass marker (i.e. a marker of a mass that does not correspond to the mass of any of the four nucleotides). In a further embodiment, the tag is a natural tag, such as a polyA tail or the natural 3' heterogeneity that can result, for example, from a transcription reaction.

Methods of sequence analysis in which nucleic acids have been replicated from a nucleic acid molecule obtained from a biological sample are specifically digested using one or more nucleases (deoxyribonucleases for DNA, and ribonucleases for RNA) are provided. The fragments captured on a solid support carrying the corresponding complementary sequences. Hybridization events and the actual molecular weights of the captured target sequences provide information on mutations in the gene. The array can be analyzed spot-by-spot using mass spectrometry. Further, the fragments generated can be ordered to provide the sequence of the larger target fragment.



-11-

In another embodiment, at least one primer with a 3'-terminal base is hybridized to the target nucleic acid near a site where possible mutations are to be detected. An appropriate polymerase and a set of three nucleoside triphosphates (NTPs) and the fourth added as a  
5 terminator are reacted. The extension reaction products are measured by mass spectrometry and are indicative of the presence and the nature of a mutation. The set of three NTPs and one dd-NTP (or three NTPs and one 3'-deoxy NTP), will be varied to be able to discriminate between several mutations (including compound heterozygotes) in the target nucleic acid  
10 sequence.

Methods for detecting and diagnosing neoplasia/malignancies in a tissue or cell sample are provided. The methods rely on a telomeric repeat amplification protocol (TRAP) -MS assay and include the steps of:

- 15 a) obtaining a tissue or a cell sample, such as a clinical isolate or culture of suspected cells;
- b) isolating/extracting/purifying telomerase from the sample;
- 20 c) adding the telomerase extract to a composition containing a synthetic DNA primer, which is optionally immobilized, complementary to the telomeric repeat, and all four dNTPs under conditions that result in telomerase specific extension of the synthetic DNA;
- 25 d) amplifying the telomerase extended DNA products,, preferably using a primer that contains a "linker moiety", such as a moiety based on thiol chemistry or streptavidin;

-12-

- e) isolating linker-amplified primers, such as by using a complementary binding partner immobilized on a solid support;
- f) optionally conditioning the DNA for crystal formation;
- 5 and
- g) performing MS by ionizing/volatizing the sample to detect the DNA product.

Telomerase-specific extension is indicative of neoplasia/malignancy.

This method can be used to detect specific malignancies. The use  
10 of MS to detect the DNA product permits identification the extended product, which is indicative of telomerase activity in the sample.

If desired, the synthetic DNA can be in the form an array.

Methods for detecting mutations are provided and the use thereof  
oncogenes and to thereby screen for transformed cells, which are  
15 indicative of neoplasia. Detection of mutations present in oncogenes are indicative of transformation. This method includes the steps of:

- a) obtaining a biological sample;
- b) amplifying a portion of the selected proto-oncogene that includes a codon indicative of transformation,  
20 where one primer has a linker moiety for immobilization;
- c) immobilizing DNA via the linker moiety to a solid support, optionally in the form of an array;
- d) hybridizing a primer complementary to the proto  
25 oncogene sequence that is upstream from the codon
- e) adding 3dNTPs/1 ddNTP and DNA polymerase and extending the hybridized primer to the next ddNTP location;

-13-

- f) ionizing/volatizing the sample; and
- g) detecting the mass of the extended DNA, whereby mass indicates the presence of wild-type or mutant alleles. The presence of a mutant allele at the codon  
5 is diagnostic for neoplasia.

In an exemplary embodiment, extension-MS analysis is used detect the presence of a mutated codon 634 in the retrovirus (RET)-proto oncogene.

In another embodiment, methods for diagnosing diseases using reverse transcription and amplification of a gene expressed in  
10 transformed cells. In particular, a method for diagnosis of neuroblastoma using reverse transcriptase (RT)-MS of tyrosine hydroxylase, which is a catecholamine biosynthetic enzyme that expressed in tumor cells, but not in tumor cells but not normal cells, such as normal bone marrow cells is provided. The method includes the steps  
15 of:

- a) obtaining a tissue sample;
- b) isolating polyA RNA from the sample;
- c) preparing a cDNA library using reverse transcription;
- d) amplifying a cDNA product, or portion thereof, of the  
20 selected gene, where one oligo primer has a linker moiety;
- e) isolating the amplified product by immobilizing the DNA to solid support via the linker moiety;
- f) optionally conditioning the DNA:
- g) ionizing/volatizing sample and detecting the presence  
25 of a DNA peak that is indicative of expression of the selected gene gene.

For example, expression of the tyrosin hydroxylase gene is indicative of neuroblastoma.

-14-

Also provided are methods of directly detecting a double-stranded nucleic acid using MALDI-TOF MS. These methods include the steps of:

- 5 a) isolating a double stranded DNA of an appropriate size for MS via amplification methods or formed by hybridization of single-stranded DNA fragment;
- b) preparing the double-stranded DNA for analysis under conditions that increase the ratio of dsDNA:ssDNA in which the conditions include one or all of the following: preparing samples for analysis at reduced temperatures (i.e. 4 ° C),  
10 and using of higher DNA concentrations in the matrix to drive duplex formation
- c) ionizing/volatizing the sample of step b), where this step uses low acceleration voltage of the ions to assist in maintaining duplex DNA by, for example, adjusting laser  
15 power to just above threshold irradiation for ionization, and
- d) detecting the presence of the dsDNA of the appropriate mass.

In preferred embodiments, the matrix includes 3-hydroxypicolinic acid.

The detected DNA can be indicative of a genetic disorder, genetic  
20 disease, genetic predisposition to a disease chromosomal abnormalities. In other embodiments, the mass of the double stranded DNA is indicative of the deletion, insertion, mutation.

A method designated primer oligo base extension (PROBE) is provided. This method uses a single detection primer followed by an  
25 oligonucleotide extension step to give products, which can be readily resolved by MALDI-TOF mass spectrometry. The products differ in length by a number of bases specific for a number of repeat units or for second site mutations within the repeated region. The method is exemplified using as a model system the AluVpA polymorphism in intron

-15-

5 of the interferon- $\alpha$  receptor gene located on human chromosome 21, and the poly T tract of the splice acceptor site of intron 8 from the CFTR gene located on human chromosome 7. The method is advantageously used for example, for determining identity, identifying mutations, familial  
5 relationship, HLA compatibility and other such markers, using PROBE-MS analysis of microsatellite DNA. In a preferred embodiment, the method includes the steps of:

- a) obtaining a biological sample from two individuals;
- b) amplifying a region of DNA from each individual that  
10 contains two or more microsatellite DNA repeat sequences
- c) ionizing/volatizing the amplified DNA;
- d) detecting the presence of the amplified DNA and comparing the molecular weight of the amplified DNA. Different sizes are indicative of non-identity (i.e. wild-type versus  
15 mutation), non-heredity or non-compatibility; similar size fragments indicate the possibility identity, of familial relationship, or HLA compatibility.

More than one marker may be examined simultaneously, primers with different linker moieties are used for immobilization.

20 Another method loop-primer oligo base extension, designated LOOP-PROBE, for detection of mutations especially predominant disease causing mutations or common polymorphisms is provided. In a particular embodiment, this method for detecting target nucleic acid in a sample, includes the steps of:

- 25 a) amplifying a target nucleic acid sequence, such as  $\beta$ -globin, in a sample, using (i) a first primer whose 5'-end shares identity to a portion of the target DNA immediately downstream from the targeted codon followed by a sequence that introduces a unique restriction endonuclease

-16-

- site, such as Cfol in the case of  $\beta$ -globin, into the amplicon and whose 3'-end primer is self-complementary; and (ii) a second downstream primer that contains a tag, such as biotin, for immobilizing the DNA to a solid support, such as streptavidin beads;
- 5 c) immobilizing the double-stranded amplified DNA to a solid support via the linker moiety;
- d) denaturing the immobilized DNA and isolating the non-immobilized DNA strand;
- 10 e) annealing the intracomplementary sequences in the 3'-end of the isolated non-immobilized DNA strand, such that the 3'-end is extendable by a polymerase, which annealing can be performed, for example, by heating then and cooling to about 37° C, or other suitable method;
- 15 f) extending the annealed DNA by adding DNA polymerase, 3 dNTPs/1 ddNTP, whereby the 3'-end of the DNA strand is extended by the DNA polymerase to the position of the next ddNTP location (i.e., to the mutation location);
- g) cleaving the extended double stranded stem loop DNA with the unique restriction endonuclease and removing the
- 20 cleaved stem loop DNA
- i) (optionally adding a matrix) ionizing/volatizing the extended product; and
- j) detecting the presence of the extended target nucleic acid, whereby the presence of a DNA fragment of a mass
- 25 different from wild-type is indicative of a mutation at the target codon(s).

This method eliminates one specific reagent for mutation detection compared other methods of MS mutational analyses, thereby simplifying

-17-

the process and rendering it amenable to automation. Also, the specific extended product that is analyzed is cleaved from the primer and is therefore shorter compared to the other methods. In addition, the annealing efficiency is higher compared to annealing of an added primer  
5 and should therefore generate more product. The process is compatible with multiplexing and various detection schemes (e.g., single base extension, oligo base extension and sequencing). For example, the extension of the loop-primer can be used for generation of short diagnostic sequencing ladders within highly polymorphic regions to  
10 perform, for example, HLA typing or resistance as well as species typing.

In another embodiment, a method of detecting a target nucleic acid in a biological sample using RNA amplification is provided. In the method, the target is amplified the target nucleic acid, using a primer that shares a region complementary to the target sequence and upstream  
15 encodes a promoter, such as the T7 promoter. A DNA-dependent RNA polymerase and appropriate ribonucleotides are added to synthesize RNA, which is analyzed by MS.

Improved methods of sequencing DNA using MS are provided. In these methods thermocycling for amplification is used prior to MS  
20 analysis, thereby increasing the signal.

Also provided are primers for use in MS analyses. In particular, primers, comprising all or, for longer oligonucleotides, at least about 20, preferably about 16, bases of any of the sequence of nucleotides sequences set forth in SEQ ID NOs. 1-22, 24, 27-38, 41-86, 89, 92, 95,  
25 98, 101-110, 112-123, 126, 128, 129, and primers set forth in SEQ ID Nos. 280-287. The primers are unlabeled, and optionally include a mass modifying moiety, which is preferably attached to the 5' end.

Other features and advantages of the methods provided herein will be further described with reference to the following Figures, Detailed Description and Claims.

#### BRIEF DESCRIPTION OF THE FIGURES

5       FIGURE 1A is a diagram showing a process for performing mass spectrometric analysis on one target detection site (TDS) contained within a target nucleic acid molecule (T), which has been obtained from a biological sample. A specific capture sequence (C) is attached to a solid support (SS) via a spacer (S). The capture sequence is chosen to  
10       specifically hybridize with a complementary sequence on the target nucleic acid molecule (T), known as the target capture site (TCS). The spacer (S) facilitates unhindered hybridization. A detector nucleic acid sequence (D), which is complementary to the TDS is then contacted with the TDS. Hybridization between D and the TDS can be detected by mass  
15       spectrometry.

FIGURE 1B is a diagram showing a process for performing mass spectrometric analysis on at least one target detection site (here TDS 1 and TDS 2) via direct linkage to a solid support. The target sequence (T) containing the target detection site (TDS 1 and TDS 2) is immobilized to  
20       a solid support via the formation of a reversible or irreversible bond formed between an appropriate functionality (L') on the target nucleic acid molecule (T) and an appropriate functionality (L) on the solid support. Detector nucleic acid sequences (here D1 and D2), which are complementary to a target detection site (TDS 1 or TDS 2) are then  
25       contacted with the TDS. Hybridization between TDS 1 and D1 and/or TDS 2 and D2 can be detected and distinguished based on molecular weight differences.

FIGURE 1C is a diagram showing a process for detecting a wildtype ( $D^{wt}$ ) and/or a mutant ( $D^{mut}$ ) sequence in a target (T) nucleic acid



-19-

molecule. As in Figure 1A, a specific capture sequence (C) is attached to a solid support (SS) via a spacer (S). In addition, the capture sequence is chosen to specifically interact with a complementary sequence on the target sequence (T), the target capture site (TCS) to be detected through hybridization. If the target detection site (TDS) includes a mutation, X, detection sites can be distinguished from wildtype by mass spectrometry. Preferably, the detector nucleic acid molecule (D) is designed so that the mutation is in the middle of the molecule and therefore would not lead to a stable hybrid if the wildtype detector oligonucleotide ( $D^{wt}$ ) is contacted with the target detector sequence, e.g., as a control. The mutation can also be detected if the mutated detector oligonucleotide ( $D^{mut}$ ) with the matching base at the mutated position is used for hybridization. If a nucleic acid molecule obtained from a biological sample is heterozygous for the particular sequence (i.e. contain  $D^{wt}$  and  $D^{mut}$ ),  $D^{wt}$  and  $D^{mut}$  will be bound to the app and  $D^{mut}$  to be detected simultaneously.

FIGURE 2 is a diagram showing a process in which several mutations are simultaneously detected on one target sequence molecular weight differences between the detector oligonucleotides D1, D2 and D3 must be large enough so that simultaneous detection (multiplexing) is possible. This can be achieved either by the sequence itself (composition or length) or by the introduction of mass-modifying functionalities M1-M3 into the detector oligonucleotide.

FIGURE 3 is a diagram showing still another multiplex detection format. In this embodiment, differentiation is accomplished by employing different specific capture sequences which are position-specifically immobilized on a flat surface (e.g., a 'chip array'). If different target sequences T1-Tn are present, their target capture sites TCS1-TCSn will interact with complementary immobilized capture sequences C1-Cn. Detection is achieved by employing appropriately mass differentiated

-20-

detector oligonucleotides D1-Dn, which are mass differentiated either by their sequences or by mass modifying functionalities M1-Mn.

FIGURE 4 is a diagram showing a format wherein a predesigned target capture site (TCS) is incorporated into the target sequence using nucleic acid (i.e., PCR) amplification. Only one strand is captured, the other is removed (e.g., based on the interaction between biotin and streptavidin coated magnetic beads). If the biotin is attached to primer 1 the other strand can be appropriately marked by a TCS. Detection is as described above through the interaction of a specific detector oligonucleotide D with the corresponding target detection site TDS via mass spectrometry.

FIGURE 5 is a diagram showing how amplification (here ligase chain reaction (LCR)) products can be prepared and detected by mass spectrometry. Mass differentiation can be achieved by the mass modifying functionalities (M1 and M2) attached to primers (P1 and P4 respectively). Detection by mass spectrometry can be accomplished directly (*i.e.* without employing immobilization and target capturing sites (TCS)). Multiple LCR reactions can be performed in parallel by providing an ordered array of capturing sequences (C). This format allows separation of the ligation products and spot by spot identification via mass spectrometry or multiplexing if mass differentiation is sufficient.

FIGURE 6A is a diagram showing mass spectrometric analysis of a nucleic acid molecule, which has been amplified by a transcription amplification procedure. An RNA sequence is captured via its TCS sequence, so that wildtype and mutated target detection sites can be detected as above by employing appropriate detector oligonucleotides (D).

-21-

FIGURE 6B is a diagram showing multiplexing to detect two different (mutated) sites on the same RNA in a simultaneous fashion using mass-modified detector oligonucleotides M1-D1 and M2-D2.

FIGURE 6C is a diagram of a different multiplexing procedure for  
5 detection of specific mutations by employing mass modified dideoxynucleoside or 3'-deoxynucleoside triphosphates and an RNA dependent DNA polymerase. Alternatively, DNA dependent RNA polymerase and ribonucleotide phosphates can be employed. This format allows for simultaneous detection of all four base possibilities at the site  
10 of a mutation (X).

FIGURE 7A is a diagram showing a process for performing mass spectrometric analysis on one target detection site (TDS) contained within a target nucleic acid molecule (T), which has been obtained from a biological sample. A specific capture sequence (C) is attached to a solid  
15 support (SS) via a spacer (S). The capture sequence is chosen to specifically hybridize with a complementary sequence on T known as the target capture site (TCS). A nucleic acid molecule that is complementary to a portion of the TDS is hybridized to the TDS 5' of the site of a mutation (X) within the TDS. The addition of a complete set of  
20 dideoxynucleosides or 3'-deoxynucleoside triphosphates (e.g., pppAdd, pppTdd, pppCdd and pppGdd) and a DNA dependent DNA or RNA polymerase allows for the addition only of the one dideoxynucleoside or 3'-deoxynucleoside triphosphate that is complementary to X.

FIGURE 7B is a diagram showing a process for performing mass  
25 spectrometric analysis to determine the presence of a mutation at a potential mutation site (M) within a nucleic acid molecule. This format allows for simultaneous analysis of alleles (A) and (B) of a double stranded target nucleic acid molecule, so that a diagnosis of homozygous normal, homozygous mutant or heterozygous can be provided. Allele A

-22-

and B are each hybridized with complementary oligonucleotides ((C) and (D) respectively), that hybridize to A and B within a region that includes M. Each heteroduplex is then contacted with a single strand specific endonuclease, so that a mismatch at M, indicating the presence of a mutation, results in the cleavage of (C) and/or (D), which can then be detected by mass spectrometry.

FIGURE 8 is a diagram showing how both strands of a target DNA can be prepared for detection using transcription vectors having two different promoters at opposite locations (e.g., the SP6 and T7 promoter). This format is particularly useful for detecting heterozygous target detection sites (TDS). Employing the SP6 or the T7 RNA polymerase both strands could be transcribed separately or simultaneously. The transcribed RNA molecules can be specifically captured and simultaneously detected using appropriately mass-differentiated detector oligonucleotides. This can be accomplished either directly in solution or by parallel processing of many target sequences on an ordered array of specifically immobilized capturing sequences.

FIGURE 9 is a diagram showing how RNA prepared as described in Figures 6, 7 and 8 can be specifically digested using one or more ribonucleases and the fragments captured on a solid support carrying the corresponding complementary sequences. Hybridization events and the actual molecular weights of the captured target sequences provide information on whether and where mutations in the gene are present. The array can be analyzed spot by spot using mass spectrometry. DNA can be similarly digested using a cocktail of nucleases including restriction endonucleases. Mutations can be detected by different molecular weights of specific, individual fragments compared to the molecular weights of the wildtype fragments.

FIGURE 10A shows UV spectra resulting from the experiment described in the following Example 1. Panel i) shows the absorbance of the 26-mer before hybridization. Panel ii) shows the filtrate of the centrifugation after hybridization. Panel iii) shows the results after the first wash with 50 mM ammonium citrate. Panel iv) shows the results after the second wash with 50 mM ammonium citrate.

FIGURE 10B shows a mass spectrum resulting from the experiment described in the following Example 1 after three washing/centrifugation steps.

10        FIGURE 10C shows a mass spectrum resulting from the experiment described in the following Example 1 showing the successful desorption of the hybridized 26-mer off of beads in accordance with the format depicted schematically in Figure 1B.

15        FIGURE 11 shows a mass spectrum resulting from the experiment described in the following Example 1 showing the giving proof of an experiment as schematically depicted in FIGURE 1B successful desorption of the hybridized 40-mer. The efficiency of detection suggests that fragments much longer than 40-mers can also be desorbed. Figure 12 shows a mass spectrum resulting from the experiment described in the following Example 2 showing the successful desorption and differentiation of an 18-mer and 19-mer by electrospray mass spectrometry, the mixture (top), peaks resulting from 18-mer emphasized (middle) and peaks resulting from 19-mer emphasized (bottom)

20        FIGURE 13 is a graphic representation of the process for detecting the Cystic Fibrosis mutation  $\Delta F508$  as described in Example 3.

FIGURE 14 is a mass spectrum of the DNA extension product of a  $\Delta F508$  homozygous normal of Example 3.

FIGURE 15 is a mass spectrum of the DNA extension product of a  $\Delta F508$  heterozygous mutant of Example 3.

FIGURE 16 is a mass spectrum of the DNA extension product of a  $\Delta$ F508 homozygous normal of Example 3.

FIGURE 17 is a mass spectrum of the DNA extension product of a  $\Delta$ F508 homozygous mutant of Example 3.

5 — FIGURE 18 is a mass spectrum of the DNA extension product of a  $\Delta$ F508 heterozygous mutant of Example 3.

FIGURE 19 is a graphic representation of various processes for performing apolipoprotein E genotyping of Example 4.

FIGURE 20 shows the nucleic acid sequence of normal  
10 apolipoprotein E (encoded by the E3 allele, FIG. 20B) and other isotypes encoded by the E2 and E4 alleles (FIG. 20A).

FIGURE 21A shows a composite restriction pattern for various genotypes of apolipoprotein E using the CfoI restriction endonuclease.

FIGURE 21B shows the restriction pattern obtained in a 3.5%  
15 MetPhor Agarose Gel for various genotypes of apolipoprotein E.

FIGURE 21C shows the restriction pattern obtained in a 12% polyacrylamide gel for various genotypes of apolipoprotein E.

FIGURE 22A is a chart showing the molecular weights of the 91, 83, 72, 48 and 35 base pair fragments obtained by restriction enzyme  
20 cleavage of the E2, E3 and E4 alleles of apolipoprotein E.

FIGURE 22B is the mass spectrum of the restriction product of a homozygous E4 apolipoprotein E genotype.

FIGURE 23A is the mass spectrum of the restriction product of a homozygous E3 apolipoprotein E genotype.

25 FIGURE 23B is the mass spectrum of the restriction product of a E3/E4 apolipoprotein E genotype.

FIGURE 24 is an autoradiograph of Example 5 of a 7.5% polyacrylamide gel in which 10% (5 $\mu$ l) of each amplified sample was loaded: sample M: pBR322 *A*lul dig sted; sampl 1: HBV positive in

-25-

serological analysis; sample 2: also HBV positive; sample 3: without serological analysis but with an increased level of transaminases, indicating liver disease; sample 4: HBV negative containing HCV; sample 5: HBV posit-) negative control; (+) positive control). Staining was done  
5 with ethidium bromide.

FIGURE 25A is a mass spectrum of sample 1, which is HBV positive. The signal at 20754 Da represents the HBV related amplification product (67 nucleotides, calculated mass: 20735 Da). The mass signal at 10390 Da represents the  $[M + 2H]^{2+}$  molecule ion  
10 (calculated: 10378 Da).

FIGURE 25B is a mass spectrum of sample 3, which is HBV negative corresponding to nucleic acid (i.e., PCR), serological and dot blot based assays. The amplified product is generated only in trace amounts. Nevertheless it is unambiguously detected at 20751 Da  
15 (calculated mass: 20735 Da). The mass signal at 10397 Da represents the  $[M + 2H]^{2+}$  molecule ion (calculated: 10376 Da).

FIGURE 25C is a mass spectrum of sample 4, which is HBV negative, but HCV positive. No HBV specific signals were observed.

FIGURE 26 shows a part of the *E. coli lacI* gene with binding sites  
20 of the complementary oligonucleotides used in the ligase chain reaction (LCR) of Example 6. Here the wildtype sequence is displayed. The mutant contains a point mutation at bp 191 which is also the site of ligation (bold). The mutation is a C to T transition (G to A, respectively). This leads to a T-G mismatch with oligo B (and A-C mismatch with oligo  
25 C, respectively).

FIGURE 27 is a 7.15% polyacrylamide gel of Example 6 stained with thidium bromide. M: chain length standard (pUC19DNA, *MspI* digested). Lane 1: LCR with wildtype template. Lane 2: LCR with mutant template. Lane 3: (control) LCR without template. The ligation

product (50 bp) was only generated in the positive reaction containing wildtype template.

FIGURE 28 is an HPLC chromatogram of two pooled positive LCRs.

FIGURE 29 shows an HPLC chromatogram the same conditions but  
5 mutant template were used. The small signal of the ligation product is due to either template-free ligation of the educts or to a ligation at a (G-T, A-C) mismatch. The 'false positive' signal is significantly lower than the signal of ligation product with wildtype template depicted in Figure 28. The analysis of ligation educts leads to 'double-peaks' because two  
10 of the oligonucleotides are 5'-phosphorylated.

FIGURE 30 In (b) the complex signal pattern obtained by MALDI-TOF-MS analysis of *Pfu* DNA-ligase solution of Example 6 is depicted. In (a) a MALDI-TOF-spectrum of an unpurified LCR is shown. The mass signal 67569 Da probably represents the *Pfu* DNA ligase.

15 FIGURE 31 shows a MALDI-TOF spectrum of two pooled positive LCRs (a). The signal at 7523 Da represents unligated oligo A (calculated: 7521 Da) whereas the signal at 15449 Da represents the ligation product (calculated: 15450 Da). The signal at 3774 Da is the  $[M + 2H]^{2+}$  signal of oligo A. The signals in the mass range lower than 2000 Da are due to the matrix ions. The spectrum corresponds to lane 1 in figure 27 and the  
20 chromatogram in figure 28. In (b) a spectrum of two pooled negative LCRs (mutant template) is shown. The signal at 7517 Da represents oligo A (calculated: 7521 Da).

FIGURE 32 shows a spectrum of two pooled control reactions  
25 (with salmon sperm DNA as template). The signals in the mass range around 2000 Da are due to Tween20, only oligo A could be detected, as expected.

FIGURE 33 shows a spectrum of two pooled positive LCRs (a). The purification was done with a combination of ultrafiltration and



-27-

streptavidin DynaBeads as described in the text. The signal at 15448 Da represents the ligation product (calculated: 15450 Da). The signal at 7527 represents oligo A (calculated: 7521 Da). The signals at 3761 Da is the  $[M + 2H]^{2+}$  signal of oligo A, whereas the signal at 5140 Da is the  $[M + 3H]^{2+}$  signal of the ligation product. In (b) a spectrum of two pooled negative LCRs (without template) is shown. The signal at 7514 Da represents oligo A (calculated: 7521 Da).

FIGURE 34 is a schematic presentation of the oligo base extension of the mutation detection primer as described in Example 7, using ddTTP (A) or ddCTP (B) in the reaction mix, respectively. The theoretical mass calculation is given in parenthesis. The sequence shown is part of the exon 10 of the CFTR gene that bears the most common cystic fibrosis mutation  $\Delta F508$  and more rare mutations  $\Delta I507$  as well as Ile506Ser.

FIGURE 35 is a MALDI-TOF-MS spectrum recorded directly from precipitated oligo base extended primers for mutation detection. The spectrum in (A) and (B), respectively show the annealed primer (CF508) without further extension reaction. Panel C displays the MALDI-TOF spectrum of the wild type by using pppTdd in the extension reaction and D a heterozygotic extension products carrying the 506S mutation when using pppCdd as terminator. Panels E and F show a heterozygote with  $\Delta F508$  mutation with pppTdd and pppCdd as terminators in the extension reaction. Panels G and H represent a homozygous  $\Delta F508$  mutation with either pppTdd or pppCdd as terminators. The template of diagnosis is pointed out below each spectrum and the observed/expected molecular mass are written in parenthesis.

FIGURE 36 shows the portion of the sequence of pRFc1 DNA, which was used as template for nucleic acid amplification in Example 8 of unmodified and 7-deazapurine containing 99-mer and 200-mer nucleic

-28-

acids as well as the sequences of the 19-mer forward primer and the two 18-mer reverse primers.

FIGURE 37 shows the portion of the nucleotide sequence of M13mp18 RFI DNA, which was used in Example 8 for nucleic acid amplification of unmodified and 7-deazapurine containing 103-mer nucleic acids. Also shown are nucleotide sequences of the 17-mer primers used in the nucleic acid amplification reaction.

FIGURE 38 shows the result of a polyacrylamide gel electrophoresis of amplified products described in Example 8 purified and concentrated for MALDI-TOF MS analysis. M: chain length marker, lane 1: 7-deazapurine containing 99-mer amplified product, lane 2: unmodified 99-mer, lane 3: 7-deazapurine containing 103-mer and lane 4: unmodified 103-mer amplified product.

FIGURE 39: an autoradiogram of polyacrylamide gel electrophoresis of nucleic acid (i.e., PCR) reactions carried out with 5'-[<sup>32</sup>P]-labeled primers 1 and 4. Lanes 1 and 2: unmodified and 7-deazapurine modified 103-mer amplified product (53321 and 23520 counts), lanes 3 and 4: unmodified and 7-deazapurine modified 200-mer (71123 and 39582 counts) and lanes 5 and 6: unmodified and 7-deazapurine modified 99-mer (173216 and 94400 counts).

FIGURE 40 a) MALDI-TOF mass spectrum of the unmodified 103-mer amplified products (sum of twelve single shot spectra). The mean value of the masses calculated for the two single strands (31768 u and 31759 u) is 31763 u. Mass resolution: 18. b) MALDI-TOF mass spectrum of 7-deazapurine containing 103-mer amplified product (sum of three single shot spectra). The mean value of the masses calculated for the two single strands (31727 u and 31719 u) is 31723 u. Mass resolution: 67.

FIGURE 41: a) MALDI-TOF mass spectrum of the unmodified 99-mer amplified product (sum of twenty single shot spectra). Values of the masses calculated for the two single strands: 30261 u and 30794 u. b) MALDI-TOF mass spectrum of 7-deazapurine containing 99-mer amplified product (sum of twelve single shot spectra). Values of the masses calculated for the two single strands: 30224 u and 30750 u.

FIGURE 42: a) MALDI-TOF mass spectrum of the unmodified 200-mer amplified product (sum of 30 single shot spectra). The mean value of the masses calculated for the two single strands (61873 u and 61595 u) is 61734 u. Mass resolution: 28. b) MALDI-TOF mass spectrum of 7-deazapurine containing 200-mer amplified product (sum of 30 single shot spectra). The mean value of the masses calculated for the two single strands (61772 u and 61714 u) is 61643 u. Mass resolution: 39.

FIGURE 43: a) MALDI-TOF mass spectrum of 7-deazapurine containing 100-mer amplified product with ribomodified primers. The mean value of the masses calculated for the two single strands (30529 u and 31095 u) is 30812 u. b) MALDI-TOF mass spectrum of the amplified product after hydrolytic primer-cleavage. The mean value of the masses calculated for the two single strands (25104 u and 25229 u) is 25167 u. The mean value of the cleaved primers (5437 u and 5918 u) is 5677 u.

FIGURE 44 A-D shows the MALDI-TOF mass spectrum of the four sequencing ladders obtained from a 39-mer template (SEQ ID No. 23), which was immobilized to streptavidin beads via a 3' biotinylation. A 14-mer primer (SEQ ID NO. 24) was used in the sequencing according to Example 9.

FIGURE 45 shows a MALDI-TOF mass spectrum of a solid phase sequencing of a 78-mer template (SEQ ID No. 25), which was

-30-

immobilized to streptavidin beads via a 3' biotinylation. A 18-mer primer (SEQ ID No. 26) and ddGTP were used in the sequencing.

FIGURE 46 shows a scheme in which duplex DNA probes with single-stranded overhang capture specific DNA templates and also serve as primers for solid phase sequencing.

FIGURE 47 A-D shows MALDI-TOF mass spectra obtained from a sequencing reaction using 5' fluorescent labeled 23-mer (SEQ ID No. 29) annealed to a 3' biotinylated 18-mer (SEQ ID No. 30), leaving a 5-base overhang, which captured a 15-mer template (SEQ ID No. 31) as described in Example 9.

FIGURE 48 shows a stacking fluorogram of the same products obtained from the reaction described in FIGURE 47, but run on a conventional DNA sequencer.

FIGURE 49 shows a MALDI-TOF mass spectrum of the sequencing ladder using cycle sequencing as described in Example 1 generated from a biological amplified product as template and a 12mer (5'-TGC ACC TGA CTC-3' (SEQ ID NO. 34)) sequencing primer. The peaks resulting from depurinations and peaks which are not related to the sequence are marked by an asterisk. MALDI-TOF MS measurements were taken on a reflectron TOF MS. A.) Sequencing ladder stopped with ddATP; B.) Sequencing ladder stopped with ddCTP; C.) Sequencing ladder stopped with ddGTP; D.) Sequencing ladder stopped with ddTTP.

FIGURE 50 shows a schematic representation of the sequencing ladder generated in Fig. 49 with the corresponding calculated molecular masses up to 40 bases after the primer. For the calculation, the following masses were used: 3581.4Da for the primer, 312.2 Da for 7-deaza-dATP, 304.2 Da for dTTP, 289.2 Da for dCTP and 328.2 Da for 7-deaza-dGTP.

-31-

FIGURE 51 shows the sequence of the amplified 209 bp amplified product within the  $\beta$ -globin gene, which was used as a template for sequencing. The sequences of the appropriate amplification primer and the location of the 12mer sequencing primer is also shown. This

5 sequence represents a homozygote mutant at the position 4 bases after the primer. In a wildtype sequence this T would be replaced by an A.

FIGURE 52 shows a sequence which is part of the intron 5 of the interferon-receptor gene that bears the AluVpA polymorphism as further described in Example 11. The scheme presents the primer oligo base  
10 extension (PROBE) using ddGTP, ddCTP, or both for termination, respectively. The polymorphism detection primer (IFN) is underlined, the termination nucleotides are marked in bold letters. The theoretical mass values from the alleles found in 28 unrelated individuals and a five member family are given in the table. Both second site mutations found  
15 in most 13 units allele, but not all, are indicated.

FIGURE 53 shows the MALDI-TOF-MS spectra recorded directly from precipitated extended cyclePROBE reaction products. Family study using AluVpA polymorphism in intron 5 of the interferon- $\alpha$  receptor gene (Example 11).

20 FIGURE 54 shows the mass spectra from PROBE products using ddC as termination nucleotide in the reaction mix. The allele with the molecular mass of approximately 11650 da from the DNA of the mother and child 2 is a hint to a second site mutation within one of the repeat units.

25 FIGURE 55 shows a schematic presentation of the PROBE method for detection of different alleles in the polyT tract at the 3'-end of intron 8 of the CFTR gene with pppCdd as terminator (Example 11).

FIGURE 56 shows the MALDI-TOF-MS spectra recorded directly from the precipitated extended PROBE reaction products. Detection of all

three common alleles of the polyT tract at the 3' end of Intron 8 of the CFTR gene. (a) T5/T9 heterozygous, (b) T7/T9 heterozygous (Example 11).

FIGURE 57 shows a mass spectrum of the digestion of a 252-mer ApoE gene amplified product ( $\epsilon 3/\epsilon 3$  genotype) as described in Example 12 using a) CfoI alone and b) CfoI plus RsaI. Asterisks: depurination peaks.

FIGURE 58 shows a mass spectrum of the ApoE gene amplified product ( $\epsilon 3/\epsilon 3$  genotype) digested by CfoI and purified by a) single and b) double ethanol/glycogen and c) double isopropyl alcohol/glycogen precipitations.

FIGURE 59 shows a mass spectrum of the CfoI/RsaI digest products from a)  $\epsilon 2/\epsilon 3$ , b)  $\epsilon 3/\epsilon 3$ , c)  $\epsilon 3/\epsilon 4$ , and d)  $\epsilon 4/\epsilon 4$  genotypes. Dashed lines are drawn through diagnostic fragments.

FIGURE 60 shows a scheme for rapid identification of unknown ApoE genotypes following simultaneous digestion of a 252-mer apo E gene amplified product by the restriction enzymes CfoI and RsaI.

FIGURE 61 shows the multiplex (codons 112 and 158) mass spectrum PROBE results for a)  $\epsilon 2/\epsilon 3$ , b)  $\epsilon 3/\epsilon 3$ , c)  $\epsilon 3/\epsilon 4$ , and d)  $\epsilon 4/\epsilon 4$  genotypes. E: extension products; P: unextended primer. Top: codon 112 and 158 regions, with polymorphic sites bold and primer sequences underlined.

FIGURE 62 shows a mass spectrum of a TRAP assay to detect telomerase activity (Example 13). The spectrum shows two of the primer signals of the amplified product TS primer at 5,497.3 Da (calc. 5523 Da) and the biotinylated bioCX primer at 7,537.6 Da (calc. 7,537 Da) and the first telomerase-specific assay product containing three telomeric repeats at 12,775.8 Da (calc. 12,452 Da) its mass is larger by

-33-

one dA nucleotide (12,765 Da) due to extendase activity of Taq DNA polymerase.

FIGURE 63 depicts the higher mass range of FIGURE 62, i.e. the peak at 12,775.6 Da represents the products with these telomeric repeats. The peaks at 20,322.1 Da is the result of a telomerase activity to form seven telomeric repeats (calc. 20,395 Da including the extension by one dA nucleotide). The peaks marked 1, 2, 3 and 4 contain a four telomeric repeats at 14,674 Da as well as secondary ion product.

FIGURE 64 displays a MALDI-TOF spectrum of the RT-amplified product of the human tyrosine hydroxylase mRNA indicating the presence of neuroblastoma cells (Example 14). The signal at 18,763.8 Da represents the non-biotinylated single-stranded 61 mer of the nested amplified product (calc. 18,758.2 Da).

FIGURE 65 (a) shows a schematic representation of a PROBE reaction for the RET proto-oncogene with a mixture of dATP, dCTP, dGTP, and ddTTP (Example 15). B represents biotin, through which the sense template strand is bound through streptavidin to a solid support. Figure 65(b) shows the expected PROBE products for ddT and ddA reactions for wildtype, C→T, and C→A antisense strands.

FIGURE 66 shows the PROBE product mass spectra for (a) negative control, (b) Patient 1 being heterozygote (Wt/C→T) and (c) Patient 2 being heterozygote (Wt/C→A), reporting average  $M_i$  values.

FIGURE 67 shows the MALDI-FTMS spectra for synthetic analogs representing ribo-cleaved RET proto-oncogene amplified products from (a) wildtype, (b) G→A, and (c) G→T homozygotes, and (d) wildtype/G→A, (e) wildtype/G→T, and (f) G→A/G→T heterozygotes, reporting masses of most abundant isotope peaks.

FIGURE 68 is a schematic representation of nucleic acid immobilization via covalent bifunctional trityl linkers.

FIGURE 69 is a schematic representation of nucleic acid immobilization via hydrophobic trityl linkers.

FIGURE 70 shows a MALDI-TOF mass spectrum of a supernatant of the matrix treated Dynabeads containing bound oligo (5'-iminobiotin - TGCACCTGACTC, SEQ ID NO. 56). An internal standard (CTGTGGTCGTGC, SEQ ID NO. 57) was included in the matrix.

FIGURE 71 shows a MALDI-TOF mass spectrum of a supernatant of the matrix treated Dynabeads containing bound oligo (5'-iminobiotin - TGCACCTGACTC, SEQ ID NO. 56). An internal standard (CTGTGGTCGTGC, SEQ ID NO. 57) was included in the matrix.

FIGURE 72 schematically depicts the steps involved with the Loop-primer oligo base extension (Loop-probe) reaction.

FIGURE 73A shows a MALDI-TOF mass spectrum of a supernatant after Cfol digest of a stem loop. Figure 73B-D show MALDI-TOF mass spectrum of different genotypes: HbA the wildtype genotype (74B), HbC, a mutation of codon 6 of the  $\beta$ -globin gene which causes sickle cell disease (74C), and HbS, a different mutation of codon 6 of the  $\beta$ -globin gene which causes sickle cell disease (74D).

FIGURE 74 shows the nucleic acid sequence of the amplified region of CKR-5. The underlined sequence corresponds to the region homologous to the amplification primers. The dotted region corresponds to the 32 bp deletion.

FIGURE 75 shows the sense primer ckrT7f. Being designed to facilitate binding of T7-RNA polymerase and amplification of the CKR-5 region to be analyzed, it starts with a randomly chosen sequence of 24 bases, the T7 promoter sequence of 18 bases and the sequence homologous to CKR-5 of 19 bases.



-35-

FIGURE 76 is a MALDI-TOF mass spectrum of the CKR-5 amplification product, which was generated as described in the following Example 21.

FIGURE 77 is a positive ion UV-MALDI mass spectra of a synthetic RNA 25-mer (5'-UCCGGUCUGAUGAGUCCGUGAGGAC-3' SEQ ID NO. 62) digested with selected RNAses. For each enzyme 0.6  $\mu$ l aliquots of the 4.5  $\mu$ l assay containing a total of ca. 20 pmol of the RNA were fixed with 1.5  $\mu$ l matrix (3-HPA) for analysis. Fragments with retained 5'-terminus are marked by different arrows, specific for the different RNAses, (Hahner *et al.*, Proceedings of the 44<sup>th</sup> ASMS Conference on Mass Spectrometry and Allied Topics, p. 983 (1996)).

FIGURE 78 is an investigation of the specificity of the RNAses CL<sub>3</sub> and Cusativin by positive ion UV-MALDI mass spectra of a synthetic RNA 20-mer. Expected and/or observed cleavage sites are indicated by arrows. A, B, C indicate correct cleavage sites and corresponding singly cleaved fragments. Missing cleavages are designated by a question mark (?), unspecific cleavages by an X.

FIGURE 79 shows the separation of a mixture of DNA molecules (12-mer, 5'-biot. 19-mer, 22-mer and 5'-biot. 27-mer) with streptavidin-coated magnetic beads. a) positive ion UV-MALDI mass spectrum of 0.6  $\mu$ l of a mixture containing ca. 2-4 pmol of each species mixed with 1.5  $\mu$ l matrix (3-HPA). b) same as a) but incubation of the mixture with magnetic beads and subsequent release of the captured fragments.

FIGURE 80 Elution of immobilized 5' biotinylated 49 nt *in vitro* transcript from the streptavidin-coated magnetic beads. Positive UV-MALDI mass spectrum of the transcript prior to incubation with the magnetic beads (a). Spectra of the immobilized RNA transcript after elution with 95% formamide alone (b) and with various additives such as 10mM EDTA (c), 10mM CDTA (d) and 25% ammonium hydroxide (e);

EDTA and CDTA were adjusted with 25% ammonium hydroxide to a pH of 8.

FIGURE 81 Positive UV-MALDI mass spectra of the 5' biotinylated 49 nt in vitro transcript after RNase U<sub>2</sub> digest for 15 minutes. a)

5 Spectrum of the 25 ul assay containing ca. 100 pmol of the target RNA before separation; b) spectrum after isolation of the 5'-biotinylated fragments with magnetic beads. Captured fragments were released by a solution of 95% formamide containing 10 mM CDTA. 1 ul aliquots of the samples were mixed with 1.5 ul matrix (3-HPA) in both cases.

10 FIGURE 82 schematically depicts detection of putative mutations in the human  $\beta$ -globin gene at codon 5 and 6 and at codon 30, and the IVS-1 donor site, respectively, done in parallel. FIGURE 82A shows amplification of genomic DNA using the primers  $\beta$ 2 and  $\beta$ 11. The location of the primers and identification tags as well as an indication of  
15 the wild type and mutant sequences are shown. FIGURE 82B shows analysis of both sites in a simple Primer Reaction Oligo Base Extension (PROBE) using primers  $\beta$ -TAG1 (which binds upstream of codon 5 and 6) and  $\beta$ -TAG2 (which binds upstream of codon 30 and the IVS-1 donor site). Reaction products are captured using streptavidin-coated  
20 paramagnetic particle bound biotinylated capture primers (cap-tag-1 and cap-tag-2, respectively), that have 6 bases at the 5' end that are complementary to the 5' end of  $\beta$ -TAG1 and  $\beta$ -TAG2, respectively, and a portion which binds to a universal primer.

FIGURE 83 shows a mass spectrum of the PROBE products of a  
25 DNA sample from one individual analyzed as described schematically in FIGURE 82.

FIGURE 84 shows a mass spectrum of the sequence bound to cap-tag-2.

-37-

FIGURE 85 shows a mass spectrum obtained by using the  $\beta$ -TAG1 and  $\beta$ -TAG 2 primers in one sequencing reaction using ddATP for termination and then sorting according to the method depicted in FIGURE 82.

5        FIGURE 86 shows a mass spectrum obtained by using the  $\beta$ -TAG1 and  $\beta$ -TAG2 primers in one sequencing reaction using ddCTP for termination and then sorting according to the method depicted in FIGURE 82.

10        FIGURE 87A shows the wildtype sequence of a fragment of the chemokine receptor CKR-5 gene with primers (bold) used for amplification. The 32 base pair (bp) deletion in the CKR-5 allele is underlined; and the stop nucleotides are in italic. In FIGURE 87B, the wildtype strands are depicted with and without an added Adenosine, their length and molecular masses are indicated. FIGURE 87C indicates  
15        the same for the 32 bp deletion. FIGURE 87D shows the PROBE products for the wildtype gene and FIGURE 87E shows the mutated allele.

FIGURE 88 shows the amplification products of different unrelated individuals as analyzed by native polyacrylamide gel electrophoreses  
20        (15%) and silver stain. The band corresponding to a wildtype CKR-5 runs at 75 bp and the band from the gene with the deletion at 43 bp. Bands bigger than 75 bp are due to unspecific amplification.

FIGURE 89A shows a spectrograph of DNA derived from a heterozygous individual: the peak with a mass of 23319 Da corresponds  
25        to the wildtype CKR-5 and the peaks with masses of 13137 Da and 13451 Da to the deletion allele with and without an extra Adenosine, respectively. FIGURE 89B shows a spectrograph of DNA obtained from the same individual as in FIGURE 89A, but the DNA was treated with T4 DNA polymerase to remove the added Adenosine. FIGURES 89C and

89D are spectrographs derived from homozygous individuals and in FIGURE 89D, the Adenosine has been removed. All peaks with masses lower than 13000 Da are due to multiple charged molecules.

FIGURE 90A shows the mass spectrum of the results of a PROBE  
5 reaction performed on DNA obtained from a heterozygous individual.  
FIGURE 90B shows a mass spectrum of the results of a PROBE reaction on a homozygous individual. The peaks with masses of 6604 Da and 6607 Da, respectively correspond to the wildtype allele, and the peak with a mass of 6275 Da to the deletion allele. The primer is detected  
10 with a mass of 5673 and 5676 Da, respectively.

FIGURE 91 shows a MALDI-TOF MS spectra of a thermocycling primer Oligo Base Extension (tc-PROBE) reaction as described in Example 24 using three different templates and 5 different PROBE primers simultaneously in one reaction.

15 FIGURE 92 schematically depicts a single tube process for amplifying and sequencing exons 5-8 of the p53 gene as described in Example 25. The mass spectrum is the A reaction of Figure 93.

FIGURE 93 shows a superposition plot of four separate reactions for sequencing a portion of exon 7 of the p53 gene as described in  
20 Example 25.

FIGURE 94 shows the mass spectrum obtained from the A reaction for sequencing a portion of exon 7 of the p53 gene as described in Example 25.

FIGURE 95 shows the mass spectrum of a p53 sequencing ladder  
25 for which 5nL of each reaction were transferred to wells of a chip and measured by MALDI-TOF.

FIGURE 96A shows a MALDI-TOF mass spectra of a synthetic 50-mer (15.34 kDa) mixed with 27-mer<sub>nc</sub> (non-complementary, 8.30 kDa).

-39-

FIGURE 96B shows a MALDI-TOF mass spectra of a synthetic 50-mer (15.34k Da) mixed with a 27-mer<sub>c</sub> (complementary, 8.34 kDa). The final concentration of each oligonucleotide was 10 $\mu$ M. The signal at 23.68 kDa in Figure 96B corresponds to WC-specific dsDNA.

5     FIGURE 97A shows a MALDI-TOF mass spectrum of CfoI/RsaI digest products of a region of exon 4 of the apolipoprotein E gene ( $\epsilon$ 3 genotype), using sample preparation as in Figure 96.

FIGURE 97B is the same as Figure 97A, except with samples prepared for MALDI-TOF analysis at 4°C.

10     FIGURE 98 shows a MALDI-TOF mass spectrum of CfoI/RsaI simultaneously double digest products of a 252 base pair region of exon 4 of the apolipoprotein E gene ( $\epsilon$ 4 genotype), with samples prepared at 4°C.

15     FIGURE 99 shows the mass spectra obtained on a small population study of 15 patients with a 16 element array of diagnostic products transferred to a MALDI target using a pintool microdispenser.

FIGURE 100 is a MALDI mass spectrum of an aliquot sampled after a T<sub>1</sub> digest of a synthetic 20-mer RNA.

#### DETAILED DESCRIPTION OF THE INVENTION

##### 20     Definitions

Unless defined otherwise, all technical and scientific terms used herein have the same meaning as is commonly understood by one of skill in the art to which this invention belongs. Where permitted the subject matter of each of the co-pending patent applications and the patent is  
25     herein incorporated in its entirety.

As used herein, the term "biological sample" refers to any material obtained from any living source (e.g., human, animal, plant, bacteria, fungi, protist, virus). For purposes herein, the biological sample will typically contain a nucleic acid molecule. Examples of appropriate

-40-

biological samples include, but are not limited to: solid materials (e.g., tissue, cell pellets, biopsies) and biological fluids (e.g., urine, blood, saliva, amniotic fluid, mouth wash, cerebral spinal fluid and other body fluids).

- 5 As used herein, the phrases "chain-elongating nucleotides" and "chain-terminating nucleotides" are used in accordance with their art recognized meaning. For example, for DNA, chain-elongating nucleotides include 2'-deoxyribonucleotides (e.g., dATP, dCTP, dGTP and dTTP) and chain-terminating nucleotides include 2', 3'-dideoxyribonucleotides (e.g.,  
10 ddATP, ddCTP, ddGTP, ddTTP). For RNA, chain-elongating nucleotides include ribonucleotides (e.g., ATP, CTP, GTP and UTP) and chain-terminating nucleotides include 3'-deoxyribonucleotides (e.g., 3'dA, 3'dC, 3'dG and 3'dU). A complete set of chain elongating nucleotides refers to dATP, dCTP, dGTP and dTTP. The term "nucleotide" is also  
15 well known in the art.

- As used herein, nucleotides include nucleoside mono-, di-, and triphosphates. Nucleotides also include modified nucleotides such as phosphorothioate nucleotides and deazapurine nucleotides. A complete set of chain-elongating nucleotides refers to four different nucleotides  
20 that can hybridize to each of the four different bases comprising the DNA template.

- As used herein, the superscript 0-i designates  $i + 1$  mass differentiated nucleotides, primers or tags. In some instances, the superscript 0 can designate an unmodified species of a particular  
25 reactant, and the superscript  $i$  can designate the  $i$ -th mass-modified species of that reactant. If, for example, more than one species of nucleic acids are to be concurrently detected, then  $i + 1$  different mass-modified detector oligonucleotides ( $D^0, D^1, \dots, D^i$ ) can be used to

-41-

distinguish each species of mass modified detector oligonucleotides (D) from the others by mass spectrometry.

As used herein, "multiplexing" refers to the simultaneously detection of more than one analyte, such as more than one (mutated) loci  
5 on a particular captured nucleic acid fragment (on one spot of an array).

As used herein, the term "nucleic acid" refers to single-stranded and/or double-stranded polynucleotides such as deoxyribonucleic acid (DNA), and ribonucleic acid (RNA) as well as analogs or derivatives of either RNA or DNA. Also included in the term "nucleic acid" are analogs  
10 of nucleic acids such as peptide nucleic acid (PNA), phosphorothioate DNA, and other such analogs and derivatives.

As used herein, the term "conjugated" refers stable attachment, preferably ionic or covalent attachment. Among preferred conjugation means are: streptavidin- or avidin- to biotin interaction; hydrophobic  
15 interaction; magnetic interaction (e.g., using functionalized magnetic beads, such as DYNABEADS, which are streptavidin-coated magnetic beads sold by Dynal, Inc. Great Neck, NY and Oslo Norway); polar interactions, such as "wetting" associations between two polar surfaces or between oligo/polyethylene glycol; formation of a covalent bond, such  
20 as an amide bond, disulfide bond, thioether bond, or via crosslinking agents; and via an acid-labile or photocleavable linker.

As used herein equivalent, when referring to two sequences of nucleic acids means that the two sequences in question encode the same sequence of amino acids or equivalent proteins. When "equivalent" is  
25 used in referring to two proteins or peptides, it means that the two proteins or peptides have substantially the same amino acid sequence with only conservative amino acid substitutions that do not substantially alter the activity or function of the protein or peptid . When "equivalent" refers to a property, the property does not need to be

-42-

present to the same extent [e.g., two peptides can exhibit different rates of the same type of enzymatic activity], but the activities are preferably substantially the same. "Complementary," when referring to two nucleotide sequences, means that the two sequences of nucleotides are

5 -capable of hybridizing, preferably with less than 25%, more preferably with less than 15%, even more preferably with less than 5%, most preferably with no mismatches between opposed nucleotides. Preferably the two molecules will hybridize under conditions of high stringency.

As used herein: stringency of hybridization in determining

10 percentage mismatch are those conditions understood by those of skill in the art and typically are substantially equivalent to the following:

- 1) high stringency: 0.1 x SSPE, 0.1% SDS, 65°C
- 2) medium stringency: 0.2 x SSPE, 0.1% SDS, 50°C
- 3) low stringency: 1.0 x SSPE, 0.1% SDS, 50°C

15 It is understood that equivalent stringencies may be achieved using alternative buffers, salts and temperatures.

As used herein, a primer when set forth in the claims refers to a primer suitable for mass spectrometric methods requiring immobilizing, hybridizing, strand displacement, sequencing mass spectrometry refers to

20 a nucleic acid must be of low enough mass, typically about 70 nucleotides or less than 70, and of sufficient size to be useful in the mass spectrometric methods described herein that rely on mass spectrometric detection. These methods include primers for detection and sequencing of nucleic acids, which require a sufficient number

25 nucleotides to form a stable duplex, typically about 6-30, preferably about 10-25, more preferably about 12-20. Thus, for purposes herein a primer will be a sequence of nucleotides comprising about 6-70, more preferably a 12-70, more preferably greater than about 14 to an upper limit of 70, depending upon sequence and application of the primer. The



-43-

primers herein, for example for mutational analyses, are selected to be upstream of loci useful for diagnosis such that when performing sequencing up to or through the site of interest, the resulting fragment is of a mass that sufficient and not too large to be detected by mass spectrometry. For mass spectrometric methods, mass tags or modifier are preferably included at the 5'-end, and the primer is otherwise unlabeled.

As used herein, "conditioning" of a nucleic acid refers to modification of the phosphodiester backbone of the nucleic acid molecule (e.g., cation exchange) for the purpose of eliminating peak broadening due to a heterogeneity in the cations bound per nucleotide unit. Contacting a nucleic acid molecule with an alkylating agent such as alkyl iodide, iodoacetamide,  $\beta$ -iodoethanol, or 2,3-epoxy-1-propanol, the monothio phosphodiester bonds of a nucleic acid molecule can be transformed into a phosphotriester bond. Likewise, phosphodiester bonds may be transformed to uncharged derivatives employing trialkylsilyl chlorides. Further conditioning involves incorporating nucleotides that reduce sensitivity for depurination (fragmentation during MS) e.g., a purine analog such as N7- or N9-deazapurine nucleotides, or RNA building blocks or using oligonucleotide triesters or incorporating phosphorothioate functions that are alkylated or employing oligonucleotide mimetics such as peptide nucleic acid (PNA).

As used herein, substrate refers to an insoluble support onto which a sample is deposited according to the materials described herein. Examples of appropriate substrates include beads (e.g., silica gel, controlled pore glass, magnetic, agarose gels and crosslinked dextroses (i.e. Sepharose and Sephadex, cellulose and other materials known by those of skill in the art to serve as solid support matrices.

-44-

- For examples substrates may be formed from any or combinations of: silica gel, glass, magnet, polystyrene/1% divinylbenzene resins, such as Wang resins, which are Fmoc-amino acid-4-(hydroxymethyl)phenoxymethyl-copoly(styrene-1% divinylbenzene (DVD)) resin, chlorotriyl
- 5 (2-chlorotriylchloride copolystyrene-DVB resin) resin, Merrifield (chloromethylated copolystyrene-DVB) resin metal, plastic, cellulose, cross-linked dextrans, such as those sold under the tradename Sephadex (Pharmacia) and agarose gel, such as gels sold under the tradename Sepharose (Pharmacia), which is a hydrogen bonded polysaccharide-type
- 10 agarose gel, and other such resins and solid phase supports known to those of skill in the art. The support matrices may be in any shape or form, including, but not limited to: capillaries, flat supports such as glass fiber filters, glass surfaces, metal surfaces (steel, gold, silver, aluminum, copper and silicon), plastic materials including multiwell plates or
- 15 membranes (e.g., of polyethylene, polypropylene, polyamide, polyvinylidenedifluoride), pins (e.g., arrays of pins suitable for combinatorial synthesis or analysis or beads in pits of flat surfaces such as wafers (e.g., silicon wafers) with or without plates, and beads.

- As used herein, a selectively cleavable linker is a linker that is
- 20 cleaved under selected conditions, such as a photocleavable linker, a chemically cleavable linker and an enzymatically cleavable linker (i.e., a restriction endonuclease site or a ribonucleotide/RNase digestion). The linker is interposed between the support and immobilized DNA.

#### Isolation of nucleic acids molecules

- 25 Nucleic acid molecules can be isolated from a particular biological sample using any of a number of procedures, which are well-known in the art, the particular isolation procedure chosen being appropriate for the particular biological sample. For example, freeze-thaw and alkaline lysis procedures can be useful for obtaining nucleic acid molecules from

solid materials; heat and alkaline lysis procedures can be useful for obtaining nucleic acid molecules from urine; and proteinase K extraction can be used to obtain nucleic acid from blood (see, e.g., Rolff et al. (1994) PCR: Clinical Diagnostics and Research, Springer).

- 5 To obtain an appropriate quantity of a nucleic acid molecules on which to perform mass spectrometry, amplification may be necessary. Examples of appropriate amplification procedures for use herein include: cloning (Sambrook et al., Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory Press, 1989), polymerase chain  
10 reaction (PCR) (C.R. Newton and A. Graham, PCR, BIOS Publishers, 1994), ligase chain reaction (LCR) (see, e.g., Weidmann et al. (1994) PCR Methods Appl. Vol. 3, Pp. 57-64; F. Barany (1991) Proc. Natl. Acad. Sci. U.S.A. 88:189-93), strand displacement amplification (SDA) (see, e.g., Walker et al. (1994) Nucleic Acids Res. 22:2670-77) and  
15 variations such as RT-PCR (see, e.g., Higuchi et al. (1993) Bio/Technology 11:1026-1030), allele-specific amplification (ASA) and transcription based processes.

#### Immobilization of nucleic acid molecules to solid supports

- To facilitate mass spectrometric analysis, a nucleic acid molecule  
20 containing a nucleic acid sequence to be detected can be immobilized to an insoluble (i.e., a solid) support. Examples of appropriate solid supports include beads (e.g., silica gel, controlled pore glass, magnetic, Sephadex/Sepharose, cellulose), capillaries, flat supports such as glass fiber filters, glass surfaces, metal surfaces (steel, gold, silver, aluminum,  
25 copper and silicon), plastic materials including multiwell plates or membranes (e.g., of polyethylene, polypropylene, polyamide, polyvinylidenedifluoride), pins (e.g., arrays of pins suitable for combinatorial synthesis or analysis or beads in pits of flat surfaces such as wafers (e.g., silicon wafers) with or without filter plates.

-46-

Samples containing target nucleic acids can be transferred to solid supports by any of a variety of methods known to those of skill in the art. For example, nucleic acid samples can be transferred to individual wells of a substrate, e.g., silicon chip, manually or using a pintool microdispenser apparatus as described herein. Alternatively, a piezoelectric pipette apparatus can be used to transfer small nanoliter samples to a substrate permitting the performance of high throughput miniaturized diagnostics on a chip.

Immobilization can be accomplished, for example, based on hybridization between a capture nucleic acid sequence, which has already been immobilized to the support and a complementary nucleic acid sequence, which is also contained within the nucleic acid molecule containing the nucleic acid sequence to be detected (FIGURE 1A). So that hybridization between the complementary nucleic acid molecules is not hindered by the support, the capture nucleic acid can include an e.g., spacer region of at least about five nucleotides in length between the solid support and the capture nucleic acid sequence. The duplex formed will be cleaved under the influence of the laser pulse and desorption can be initiated. The solid support-bound nucleic acid molecule can be presented through natural oligoribo- or oligodeoxyribonucleotide as well as analogs (e.g., thio-modified phosphodiester or phosphotriester backbone) or employing oligonucleotide mimetics such as PNA analogs (see, e.g., Nielsen *et al.*, Science 254:1497 (1991)) which render the base sequence less susceptible to enzymatic degradation and -bound capture base sequence.

-47-

### Linkers

A target detection site can be directly linked to a solid support via a reversible or irreversible bond between an appropriate functionality (L') on the target nucleic acid molecule (T) and an appropriate functionality (L) on the capture molecule (FIGURE 1B). A reversible linkage can be such that it is cleaved under the conditions of mass spectrometry (i.e., a photocleavable bond such as a charge transfer complex or a labile bond being formed between relatively stable organic radicals).

Photocleavable linkers are linkers that are cleaved upon exposure to light (see, e.g., Goldmacher *et al.* (1992) Bioconj. Chem. 3:104-107), thereby releasing the targeted agent upon exposure to light. Photocleavable linkers that are cleaved upon exposure to light are known (see, e.g., Hazum *et al.* (1981) in Pept., Proc. Eur. Pept. Symp., 16th, Brunfeldt, K (Ed), pp. 105-110, which describes the use of a nitrobenzyl group as a photocleavable protective group for cysteine; Yen *et al.* (1989) Makromol. Chem 190:69-82, which describes water soluble photocleavable copolymers, including hydroxypropylmethacrylamide copolymer, glycine copolymer, fluorescein copolymer and methylrhodamine copolymer; Goldmacher *et al.* (1992) Bioconj. Chem. 3:104-107, which describes a cross-linker and reagent that undergoes photolytic degradation upon exposure to near UV light (350 nm); and Senter *et al.* (1985) Photochem. Photobiol 42:231-237, which describes nitrobenzyloxycarbonyl chloride cross linking reagents that produce photocleavable linkages), thereby releasing the targeted agent upon exposure to light. In preferred embodiments, the nucleic acid is immobilized using the photocleavable linker moiety that is cleaved during mass spectrometry. Presently preferred photocleavable linkers are set forth in the EXAMPLES.

-48-

Furthermore, the linkage can be formed with L' being a quaternary ammonium group, in which case, preferably, the surface of the solid support carries negative charges which repel the negatively charged nucleic acid backbone and thus facilitate the desorption required for analysis by a mass spectrometer. Desorption can occur either by the heat created by the laser pulse and/or, depending on L', by specific absorption of laser energy which is in resonance with the L' chromophore.

Thus, the L-L' chemistry can be of a type of disulfide bond (chemically cleavable, for example, by mercaptoethanol or dithioerythrol), a biotin/streptavidin system, a heterobifunctional derivative of a trityl ether group (see, e.g., Köster et al. (1990) "A Versatile Acid-Labile Linker for Modification of Synthetic Biomolecules," Tetrahedron Letters 31:7095) that can be cleaved under mildly acidic conditions as well as under conditions of mass spectrometry, a levulinyl group cleavable under almost neutral conditions with a hydrazinium/acetate buffer, an arginine-arginine or lysine-lysine bond cleavable by an endopeptidase enzyme like trypsin or a pyrophosphate bond cleavable by a pyrophosphatase, or a ribonucleotide bond in between the oligodeoxynucleotide sequence, which can be cleaved, for example, by a ribonuclease or alkali.

The functionalities, L and L', can also form a charge transfer complex and thereby form the temporary L-L' linkage. Since in many cases the "charge-transfer band" can be determined by UV/vis spectrometry (see, e.g., Organic Charge Transfer Complexes by R. Foster, Academic Press, 1969), the laser energy can be tuned to the corresponding energy of the charge-transfer wavelength and, thus, a specific desorption off the solid support can be initiated. Those skilled in the art will recognize that several combinations can serve this purpose

and that the donor functionality can be either on the solid support or coupled to the nucleic acid molecule to be detected or vice versa.

In yet another approach, a reversible L-L' linkage can be generated by homolytically forming relatively stable radicals. Under the influence of the laser pulse, desorption (as discussed above) as well as ionization will take place at the radical position. Those skilled in the art will recognize that other organic radicals can be selected and that, in relation to the dissociation energies needed to homolytically cleave the bond between them, a corresponding laser wavelength can be selected (see e.g.,

10 Reactive Molecules by C. Wentrup, John Wiley & Sons, 1984).

An anchoring function L' can also be incorporated into a target capturing sequence (TCS) by using appropriate primers during an amplification procedure, such as PCR (FIGURE 4), LCR (FIGURE 5) or transcription amplification (FIGURE 6A).

15 When performing exonuclease sequencing using MALDI-TOF MS, a single stranded DNA molecule immobilized via its 5-end to a solid support is unilaterally degraded with a 3'-processive exonuclease and the molecular weight of the degraded nucleotide is determined sequentially. Reverse Sanger sequencing reveals the nucleotide sequence of the

20 immobilized DNA. By adding a selectively cleavable linker, not only can the mass of the free nucleotides be determined but also, upon removal of the nucleotides by washing, the mass of the remaining fragment can be detected by MALDI-TOF upon cleaving the DNA from the solid support. Using selectively cleavable linkers, such as the photocleavable and

25 chemical cleavable linkers provided herein, this cleavage can be selected to occur during the ionization and volatilizing steps of MALDI-TOF. The same rationale applies for a 5' immobilized strand of a double stranded DNA that is degraded while in a duplex. Likewise, this also

-50-

applies when using a 5'-processive exonuclease and the DNA is immobilized through the 3'-end to the solid support.

- As noted, at least three version of immobilization are contemplated herein: 1) the target nucleic acid is amplified or obtained (the target sequence or surrounding DNA sequence must be known to make primers to amplify or isolated); 2) the primer nucleic acid is immobilized to the solid support and the target nucleic acid is hybridized thereto (this is for detecting the presence of or sequencing a target sequence in a sample); or 3) a double stranded DNA (amplified or isolated) is immobilized through linkage to one predetermined strand, the DNA is denatured to eliminate the duplex and then a high concentration of a complementary primer or DNA with identity upstream from the target site is added and a strand displacement occurs and the primer is hybridized to the immobilized strand.
- In the embodiments where the primer nucleic acid is immobilized on the solid support and the target nucleic acid is hybridized thereto, the inclusion of the cleavable linker allows the primer DNA to be immobilized at the 5'-end so that free 3'-OH is available for nucleic acid synthesis (extension) and the sequence of the "hybridized" target DNA can be determined because the hybridized template can be removed by denaturation and the extended DNA products cleaved from the solid support for MALDI-TOF MS. Similarly for 3), the immobilized DNA strand can be elongated when hybridized to the template and cleaved from the support. Thus, Sanger sequencing and primer oligo base extension (PROBE), discussed below, extension reactions can be performed using an immobilized primer of a known, upstreamn DNA sequence complementary to an invariable region of a target sequence. The nucleic acid from the person is obtained and the DNA sequence of a variable region (deletion, insertion, missense mutation that cause genetic



-51-

predisposition or diseases, or the presence of viral/bacterial or fungal DNA) not only is detected, but the actual sequence and position of the mutation is also determined.

In other cases, the target DNA must be immobilized and the primer  
5- annealed. This requires amplifying a larger DNA based on known  
sequence and then sequencing the immobilized fragments (i.e., the  
extended fragments are hybridized but not immobilized to the support as  
described above). In these cases, it is not desirable to include a linker  
because the MALDI-TOF spectrum is of the hybridized DNA; it is not  
10 necessary to cleave the immobilized template.

Any linker known to those of skill in the art for immobilizing  
nucleic acids to solid supports may be used herein to link the nucleic acid  
to a solid support. The preferred linkers herein are the selectively  
cleavable linkers, particularly those exemplified herein. Other linkers  
15 include, acid cleavable linkers, such as bismaleimideoxy propane,  
acid-labile trityl linkers.

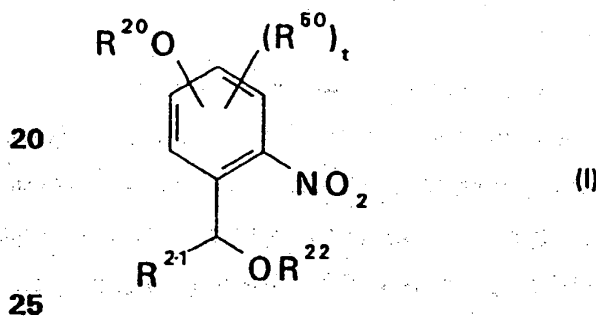
Acid cleavable linkers, photocleavable and heat sensitive linkers  
may also be used, particularly where it may be necessary to cleave the  
targeted agent to permit it to be more readily accessible to reaction.  
20 Acid cleavable linkers include, but are not limited to, bismaleimideoxy  
propane; and adipic acid dihydrazide linkers (see, e.g., Fattom et al.  
(1992) Infection & Immun. 60:584-589) and acid labile transferrin  
conjugates that contain a sufficient portion of transferrin to permit entry  
into the intracellular transferrin cycling pathway (see, e.g., Welhöner et  
25 al. (1991) J. Biol. Chem. 266:4309-4314).

-52-

### Photocleavable Linkers

Photocleavable linkers are provided. In particular, photocleavable linkers as their phosphoramidite derivatives are provided for use in solid phase synthesis of oligonucleotides. The linkers contain o-nitrobenzyl moieties and phosphate linkages which allow for complete photolytic cleavage of the conjugates within minutes upon UV irradiation. The UV wavelengths used are selected so that the irradiation will not damage the oligonucleotides and are preferably about 350-380 nm, more preferably 365 nm. The photocleavable linkers provided herein possess comparable coupling efficiency as compared to commonly used phosphoramidite monomers (see, Sinha *et al.* (1983) *Tetrahedron Lett.* 24:5843-5846; Sinha *et al.* (1984) *Nucleic Acids Res.* 12:4539-4557; Beaucage *et al.* (1993) *Tetrahedron* 49:6123-6194; and Matteucci *et al.* (1981) *J. Am. Chem. Soc.* 103:3185-3191).

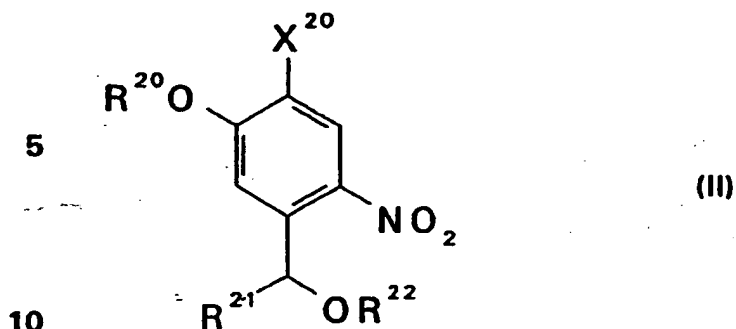
In one embodiment, the photocleavable linkers have formula I:



where  $R^{20}$  is  $\omega$ -(4,4'-dimethoxytrityloxy)alkyl or  $\omega$ -hydroxyalkyl;  $R^{21}$  is selected from hydrogen, alkyl, aryl, alkoxycarbonyl, aryloxycarbonyl and carboxy;  $R^{22}$  is hydrogen or (dialkylamino)( $\omega$ -cyanoalkoxy)P-;  $t$  is 0-3; and  $R^{60}$  is alkyl, alkoxy, aryl or aryloxy.

In a preferred embodiment, the photocleavable linkers have formula II:

-53-



where  $R^{20}$  is  $\omega$ -(4,4'-dimethoxytrityloxy)alkyl,  $\omega$ -hydroxyalkyl or alkyl;  $R^{21}$  is selected from hydrogen, alkyl, aryl, alkoxycarbonyl, aryloxycarbonyl and carboxy;  $R^{22}$  is hydrogen or (dialkylamino)( $\omega$ -cyanoalkoxy)P-; and  $X^{20}$  is hydrogen, alkyl or  $OR^{20}$ .

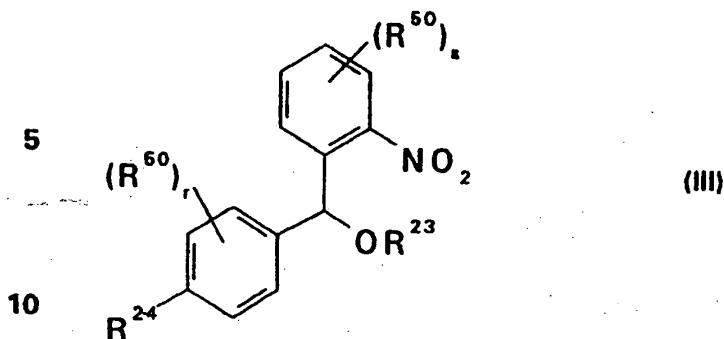
In particularly preferred embodiments,  $R^{20}$  is 3-(4,4'-dimethoxytrityloxy)propyl, 3-hydroxypropyl or methyl;  $R^{21}$  is selected from hydrogen, methyl and carboxy;  $R^{22}$  is hydrogen or (diisopropylamino)(2-cyanoethoxy)P-; and  $X^{20}$  is hydrogen, methyl or  $OR^{20}$ . In a more preferred embodiment,  $R^{20}$  is 3-(4,4'-dimethoxytrityloxy)propyl;  $R^{21}$  is methyl;  $R^{22}$  is (diisopropylamino)(2-cyanoethoxy)P-; and  $X^{20}$  is hydrogen. In another more preferred embodiment,  $R^{20}$  is methyl;  $R^{21}$  is methyl;  $R^{22}$  is (diisopropylamino)(2-cyanoethoxy)P-; and  $X^{20}$  is 3-(4,4'-dimethoxytrityloxy)propoxy.

In another embodiment, the photocleavable linkers have formula III:

30

35

-54-



where  $R^{23}$  is hydrogen or (dialkylamino)( $\omega$ -cyanoalkoxy)P-; and  $R^{24}$  is selected from  $\omega$ -hydroxyalkoxy,  $\omega$ -(4,4'-dimethoxytrityloxy)alkoxy,  $\omega$ -hydroxyalkyl and  $\omega$ -(4,4'-dimethoxytrityloxy)alkyl, and is unsubstituted or substituted on the alkyl or alkoxy chain with one or more alkyl groups;  $r$  and  $s$  are each independently 0-4; and  $R^{60}$  is alkyl, alkoxy, aryl or aryloxy. In certain embodiments,  $R^{24}$  is  $\omega$ -hydroxyalkyl or  $\omega$ -(4,4'-dimethoxytrityloxy)alkyl, and is substituted on the alkyl chain with a methyl group.

In preferred embodiments,  $R^{23}$  is hydrogen or (diisopropylamino)(2-cyanoethoxy)P-; and  $R^{24}$  is selected from 3-hydroxypropoxy, 3-(4,4'-dimethoxytrityloxy)propoxy, 4-hydroxybutyl, 3-hydroxy-1-propyl, 1-hydroxy-2-propyl, 3-hydroxy-2-methyl-1-propyl, 2-hydroxyethyl, hydroxymethyl, 4-(4,4'-dimethoxytrityloxy)butyl, 3-(4,4'-dimethoxytrityloxy)-1-propyl, 2-(4,4'-dimethoxytrityloxy)ethyl, 1-(4,4'-dimethoxytrityloxy)-2-propyl, 3-(4,4'-dimethoxytrityloxy)-2-methyl-1-propyl and 4,4'-dimethoxytrityloxymethyl.

In more preferred embodiments,  $R^{23}$  is (diisopropylamino)(2-cyanoethoxy)P-;  $r$  and  $s$  are 0; and  $R^{24}$  is selected from 3-(4,4'-dimethoxytrityloxy)propoxy, 4-(4,4'-dimethoxytrityloxy)butyl, 3-(4,4'-dimethoxytrityloxy)propyl, 2-(4,4'-dimethoxytrityloxy)ethyl, 1-(4,4'-dimethoxytrityloxy)-2-propyl, 3-(4,4'-dimethoxytrityloxy)-2-methyl-1-

-55-

propyl and 4,4'-dimethoxytrityloxymethyl.  $R^{24}$  is most preferably 3-(4,4'-dimethoxytrityloxy)propoxy.

#### Preparation of the photocleavable linkers

##### A. Preparation of photocleavable linkers of formulae I or II

5

Photocleavable linkers of formulae I or II may be prepared by the methods described below, by minor modification of the methods by choosing the appropriate starting materials or by any other methods known to those of skill in the art. Detailed procedures for the synthesis of photocleavable linkers of formula II are provided in the Examples.

10

In the photocleavable linkers of formula II where  $X^{20}$  is hydrogen, the linkers may be prepared in the following manner. Alkylation of 5-hydroxy-2-nitrobenzaldehyde with an  $\omega$ -hydroxyalkyl halide, e.g., 3-hydroxypropyl bromide, followed by protection of the resulting alcohol as, e.g., a silyl ether, provides a 5-( $\omega$ -silyloxyalkoxy)-2-nitrobenzaldehyde. Addition of an organometallic to the aldehyde affords a benzylic alcohol. Organometallics which may be used include trialkylaluminums (for linkers where  $R^{21}$  is alkyl), such as trimethylaluminum, borohydrides (for linkers where  $R^{21}$  is hydrogen), such as sodium borohydride, or metal cyanides (for linkers where  $R^{21}$  is carboxy or alkoxycarbonyl), such as potassium cyanide. In the case of the metal cyanides, the product of the reaction, a cyanohydrin, would then be hydrolyzed under either acidic or basic conditions in the presence of either water or an alcohol to afford the compounds of interest.

15

20

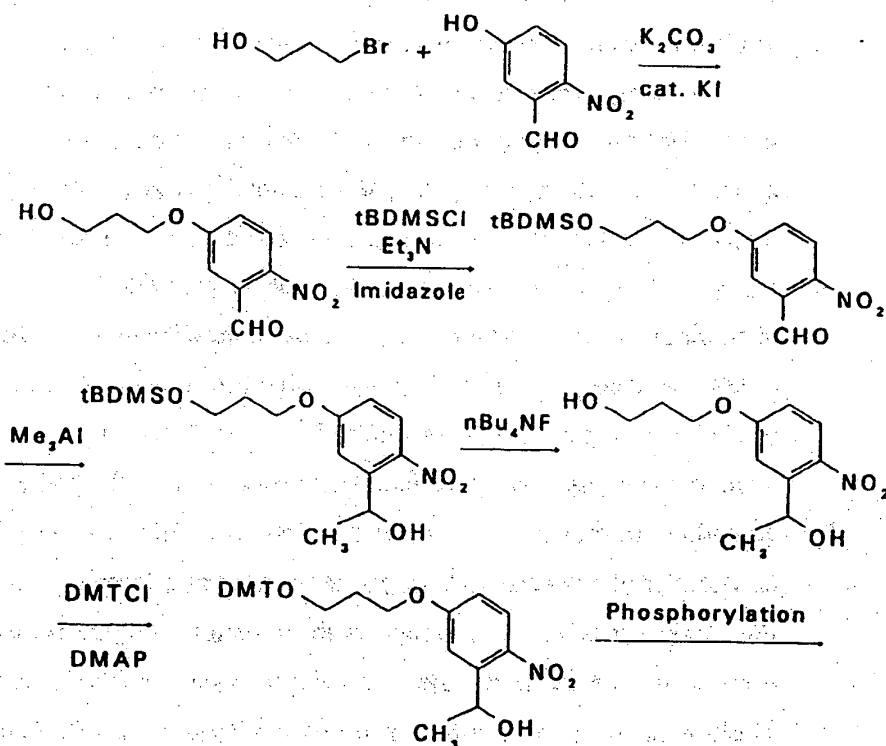
The silyl group of the side chain of the resulting benzylic alcohols may then be exchanged for a 4,4'-dimethoxytriyl group by desilylation with, e.g., tetrabutylammonium fluoride, to give the corresponding alcohol, followed by reaction with 4,4'-dimethoxytrityl chloride. Reaction

25

-56-

with, e.g., 2-cyanoethyl diisopropylchlorophosphoramidite affords the linkers where  $R^{22}$  is (dialkylamino)( $\omega$ -cyanoalkoxy)P-.

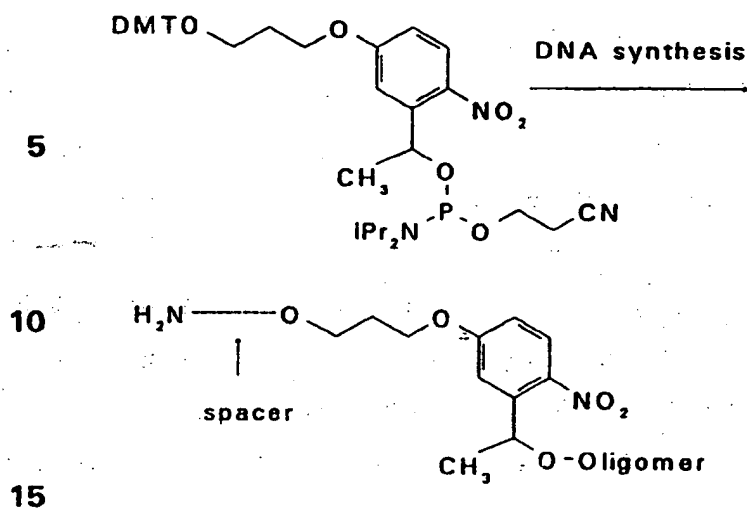
A specific example of a synthesis of a photocleavable linker of formula II is shown in the following scheme, which also demonstrates use of the linker in oligonucleotide synthesis. This scheme is intended to be illustrative only and in no way limits the scope of the invention. Experimental details of these synthetic transformations are provided in the Examples.



10

15

-57-

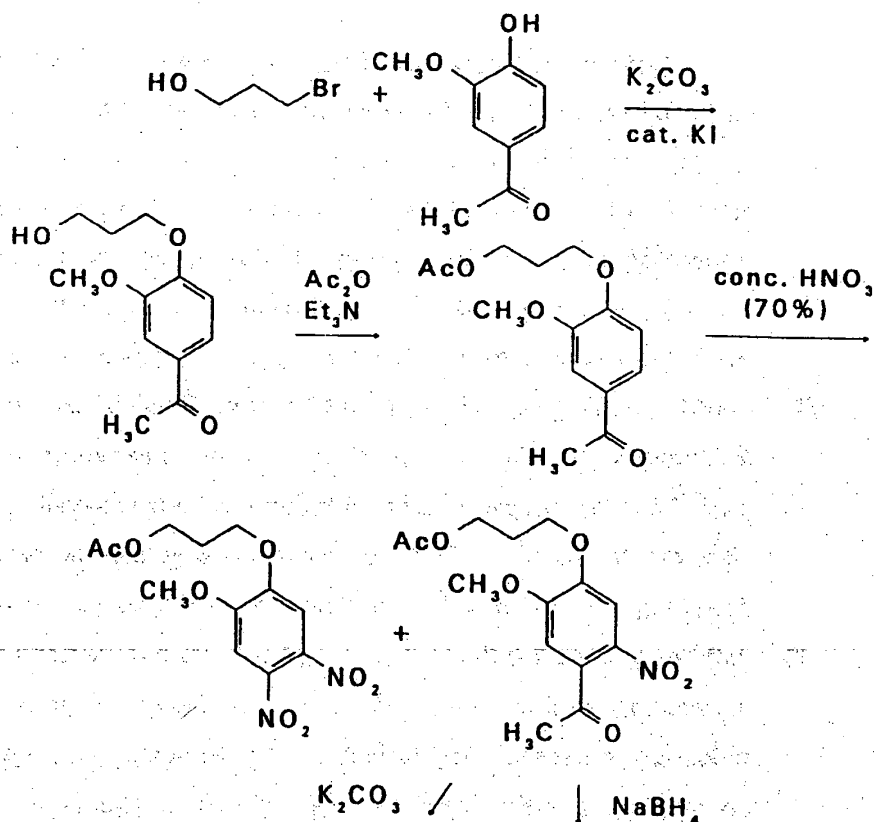


Synthesis of the linkers of formula II where X<sup>20</sup> is OR<sup>20</sup>, 3,4-dihydroxyacetophenone is protected selectively at the 4-hydroxyl by reaction with, e.g., potassium carbonate and a silyl chloride. Benzoate esters, propiophenones, butyrophenones, etc. may be used in place of the acetophenone. The resulting 4-silyloxy-3-hydroxyacetophenone is then alkylated at the 3-hydroxyl with an alkyl halide (for linkers where R<sup>20</sup> is alkyl) and desilylated with, e.g., tetrabutylammonium fluoride to afford a 3-alkoxy-4-hydroxyacetophenone. This compound is then alkylated at the 4-hydroxyl by reaction with an  $\omega$ -hydroxyalkyl halide, e.g., 3-hydroxypropyl bromide, to give a 4-( $\omega$ -hydroxyalkoxy)-3-alkoxyacetophenone. The side chain alcohol is then protected as an ester, e.g., an acetate. This compound is then nitrated at the 5-position with, e.g., concentrated nitric acid to provide the corresponding 2-nitroacetophenones. Saponification of the side chain ester with, e.g., potassium carbonate, and reduction of the ketone with, e.g., sodium borohydride, in either order gives a 2-nitro-4-( $\omega$ -hydroxyalkoxy)-5-alkoxybenzylic alcohol.

-58-

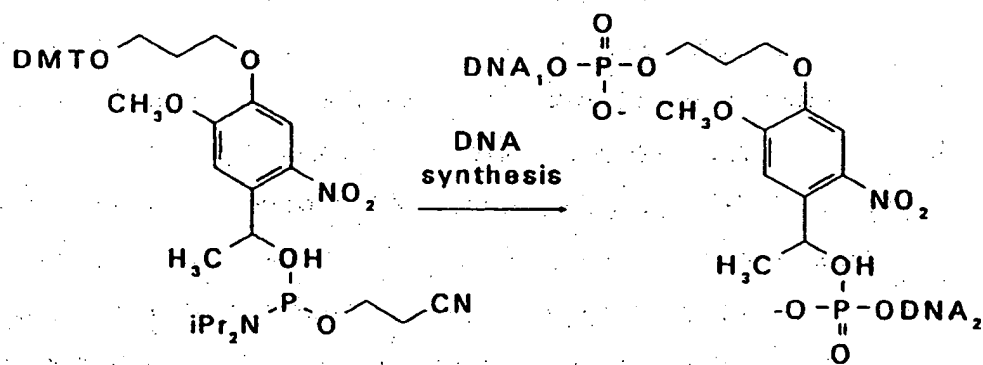
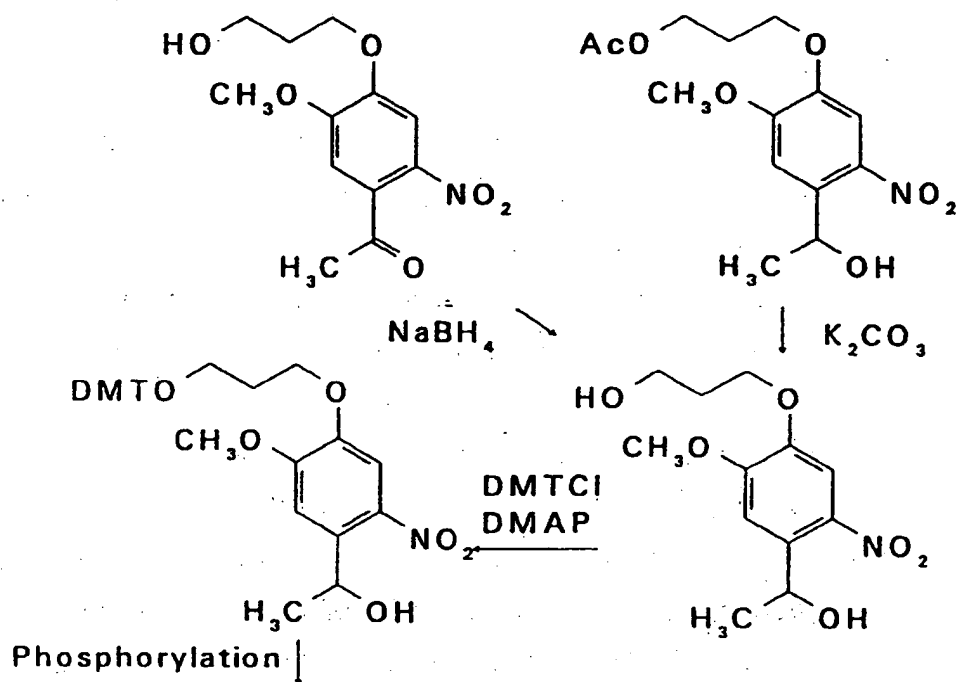
Selective protection of the side chain alcohol as the corresponding 4,4'-dimethoxytrityl ether is then accomplished by reaction with 4,4'-dimethoxytrityl chloride. Further reaction with, e.g., 2-cyanoethyl diisopropylchlorophosphoramidite affords the linkers where R<sup>22</sup> is 5-(dialkylamino)( $\omega$ -cyanoalkoxy)P-.

A specific example of the synthesis of a photocleavable linker of formula II is shown the following scheme. This scheme is intended to be illustrative only and in no way limit the scope of the invention. Detailed experimental procedures for the transformations shown are found in the 10 Examples.





-59-



-60-

### B. Preparation of photocleavable linkers of formula III

Photocleavable linkers of formula III may be prepared by the methods described below, by minor modification of the methods by choosing appropriate starting materials, or by other methods known to those of skill in the art.

In general, photocleavable linkers of formula III are prepared from  $\omega$ -hydroxyalkyl- or alkoxyaryl compounds, in particular  $\omega$ -hydroxy-alkyl or alkoxy-benzenes. These compounds are commercially available, or may be prepared from an  $\omega$ -hydroxyalkyl halide (e.g., 3-hydroxypropyl bromide) and either phenyllithium (for the  $\omega$ -hydroxyalkylbenzenes) or phenol (for the  $\omega$ -hydroxyalkoxybenzenes). Acylation of the  $\omega$ -hydroxyl group (e.g., as an acetate ester) followed by Friedel-Crafts acylation of the aromatic ring with 2-nitrobenzoyl chloride provides a 4-( $\omega$ -acetoxy-alkyl or alkoxy)-2-nitrobenzophenone. Reduction of the ketone with, e.g., sodium borohydride, and saponification of the side chain ester are performed in either order to afford a 2-nitrophenyl-4-(hydroxy-alkyl or alkoxy)phenylmethanol. Protection of the terminal hydroxyl group as the corresponding 4,4'-dimethoxytrityl ether is achieved by reaction with 4,4'-dimethoxytrityl chloride. The benzylic hydroxyl group is then reacted with, e.g., 2-cyanoethyl diisopropylchlorophosphoramidite to afford linkers of formula II where  $R^{23}$  is (dialkylamino)( $\omega$ -cyanoalkoxy)P-.

Other photocleavable linkers of formula III may be prepared by substituting 2-phenyl-1-propanol or 2-phenylmethyl-1-propanol for the  $\omega$ -hydroxy-alkyl or alkoxy-benzenes in the above synthesis. These compounds are commercially available, but may also be prepared by reaction of, e.g., phenylmagnesium bromide or benzylmagnesium bromide, with the requisite oxirane (i.e., propylene oxide) in the presence of catalytic cuprous ion.

-61-

### Chemically cleavable linkers

A variety of chemically cleavable linkers may be used to introduce a cleavable bond between the immobilized nucleic acid and the solid support. Acid-labile linkers are presently preferred chemically cleavable linkers for mass spectrometry, especially MALDI-TOF MS, because the acid labile bond is cleaved during conditioning of the nucleic acid upon addition of the 3-HPA matrix solution. The acid labile bond can be introduced as a separate linker group, e.g., the acid labile trityl groups (see Figure 68; Example 16) or may be incorporated in a synthetic nucleic acid linker by introducing one or more silyl internucleoside bridges using diisopropylsilyl, thereby forming diisopropylsilyl-linked oligonucleotide analogs. The diisopropylsilyl bridge replaces the phosphodiester bond in the DNA backbone and under mildly acidic conditions, such as 1.5% trifluoroacetic acid (TFA) or 3-HPA/1% TFA MALDI-TOF matrix solution, results in the introduction of one or more intra-strand breaks in the DNA molecule. Methods for the preparation of diisopropylsilyl-linked oligonucleotide precursors and analogs are known to those of skill in the art (see e.g., Saha et al. (1993) J. Org. Chem. 58:7827-7831). These oligonucleotide analogs may be readily prepared using solid state oligonucleotide synthesis methods using diisopropylsilyl derivatized deoxyribonucleosides.

### Nucleic acid conditioning

Prior to mass spectrometric analysis, it may be useful to "condition" nucleic acid molecules, for example to decrease the laser energy required for volatilization and/or to minimize fragmentation. Conditioning is preferably performed while a target detection site is immobilized. An example of conditioning is modification of the phosphodiester backbone of the nucleic acid molecule (e.g., cation exchange), which can be useful for eliminating peak broadening due to a

-62-

- heterogeneity in the cations bound per nucleotide unit. Contacting a nucleic acid molecule with an alkylating agent such as alkyl iodide, iodoacetamide,  $\beta$ -iodoethanol, or 2,3-epoxy-1-propanol, the monothio phosphodiester bonds of a nucleic acid molecule can be transformed into
- 5 a phosphotriester bond. Likewise, phosphodiester bonds may be transformed to uncharged derivatives employing trialkylsilyl chlorides. Further conditioning involves incorporating nucleotides that reduce sensitivity for depurination (fragmentation during MS) e.g., a purine analog such as N7- or N9-deazapurine nucleotides, or RNA building
- 10 blocks or using oligonucleotide triesters or incorporating phosphorothioate functions which are alkylated or employing oligonucleotide mimetics such as PNA.

#### Multiplex reactions

- For certain applications, it may be useful to simultaneously detect
- 15 more than one (mutated) loci on a particular captured nucleic acid fragment (on one spot of an array) or it may be useful to perform parallel processing by using oligonucleotide or oligonucleotide mimetic arrays on various solid supports. "Multiplexing" can be achieved by several different methodologies. For example, several mutations can be
- 20 simultaneously detected on one target sequence by employing corresponding detector (probe) molecules (e.g., oligonucleotides or oligonucleotide mimetics). The molecular weight differences between the detector oligonucleotides D1, D2 and D3 must be large enough so that simultaneous detection (multiplexing) is possible. This can be achieved
- 25 either by the sequence itself (composition or length) or by the introduction of mass-modifying functionalities M1-M3 into the detector oligonucleotide (see FIGURE 2).

-63-

### Mass modification of nucleic acids

Mass modifying moieties can be attached, for instance, to either the 5'-end of the oligonucleotide ( $M^1$ ), to the nucleobase (or bases) ( $M^2$ ,  $M^7$ ), to the phosphate backbone ( $M^3$ ), and to the 2'-position of the  
5 nucleoside (nucleosides) ( $M^4$ ,  $M^6$ ) and/or to the terminal 3'-position ( $M^5$ ). Examples of mass modifying moieties include, for example, a halogen, an azido, or of the type, XR, wherein X is a linking group and R is a mass-modifying functionality. The mass-modifying functionality can thus be used to introduce defined mass increments into the oligonucleotide  
10 molecule.

The mass-modifying functionality can be located at different positions within the nucleotide moiety (see, e.g., U.S. Patent No. 5,547,835 and International PCT application No. WO 94/21822). For example, the mass-modifying moiety, M, can be attached either to the  
15 nucleobase,  $M^2$  (in case of the  $c^7$ -deazanucleosides also to C-7,  $M^7$ ), to the triphosphate group at the alpha phosphate,  $M^3$ , or to the 2'-position of the sugar ring of the nucleoside triphosphate,  $M^4$  and  $M^6$ .

Modifications introduced at the phosphodiester bond ( $M^4$ ), such as with alpha-thio nucleoside triphosphates, have the advantage that these  
20 modifications do not interfere with accurate Watson-Crick base-pairing and additionally allow for the one-step post-synthetic site-specific modification of the complete nucleic acid molecule e.g., via alkylation reactions (see, e.g., Nakamaye et al. (1988) Nucl. Acids Res. 16:9947-59). Particularly preferred mass-modifying functionalities are boron-  
25 modified nucleic acids since they are better incorporated into nucleic acids by polymerases (see, e.g., Porter et al. (1995) Biochemistry 34:11963-11969; Hasan et al. (1996) Nucleic Acids Res. 24:2150-2157; Li et al. (1995) Nucl. Acids Res. 23:4495-4501).

-64-

Furthermore, the mass-modifying functionality can be added so as to affect chain termination, such as by attaching it to the 3'-position of the sugar ring in the nucleoside triphosphate, M<sup>6</sup>. For those skilled in the art, it is clear that many combinations can be used in the methods  
5 provided herein. In the same way, those skilled in the art will recognize that chain-elongating nucleoside triphosphates can also be mass-modified in a similar fashion with numerous variations and combinations in functionality and attachment positions.

Without being bound to any particular theory, the mass-  
10 modification, M, can be introduced for X in XR as well as using oligo-/polyethylene glycol derivatives for R. The mass-modifying increment in this case is 44, i.e. five different mass-modified species can be generated by just changing m from 0 to 4 thus adding mass units of 45 (m=0), 89 (m=1), 133 (m=2), 177 (m=3) and 221 (m=4) to the  
15 nucleic acid molecule (e.g., detector oligonucleotide (D) or the nucleoside triphosphates (FIGURE 6(C)), respectively). The oligo/polyethylene glycols can also be monoalkylated by a lower alkyl such as methyl, ethyl, propyl, isopropyl, t-butyl and the like. A selection of linking functionalities, X, are also illustrated. Other chemistries can be used in  
20 the mass-modified compounds (see, e.g., those described in Oligonucleotides and Analogues, A Practical Approach, F. Eckstein, editor, IRL Press, Oxford, 1991).

In yet another embodiment, various mass-modifying functionalities, R, other than oligo/polyethylene glycols, can be selected and attached via  
25 appropriate linking chemistries, X. A simple mass-modification can be achieved by substituting H for halogens like F, Cl, Br and/or I, or pseudohalogens such as CN, SCN, NCS, or by using different alkyl, aryl or aralkyl moieties such as methyl, ethyl, propyl, isopropyl, t-butyl, hexyl, phenyl, substituted phenyl, benzyl, or functional groups such as CH<sub>2</sub>F,

-65-

CHF<sub>2</sub>, CF<sub>3</sub>, Si(CH<sub>3</sub>)<sub>3</sub>, Si(CH<sub>3</sub>)<sub>2</sub>(C<sub>2</sub>H<sub>5</sub>), Si(CH<sub>3</sub>)(C<sub>2</sub>H<sub>5</sub>)<sub>2</sub>, Si(C<sub>2</sub>H<sub>5</sub>)<sub>3</sub>. Yet another mass-modification can be obtained by attaching homo- or heteropeptides through the nucleic acid molecule (e.g., detector (D)) or nucleoside triphosphates. One example, useful in generating mass-

5. modified species with a mass increment of 57, is the attachment of oligoglycines, e.g., mass-modifications of 74 (r=1, m=0), 131 (r=1, m=1), 188 (r=1, m=2), 245 (r=1, m=3) are achieved. Simple oligoamides also can be used, e.g., mass-modifications of 74 (r=1, m=0), 88 (r=2, m=0), 102 (r=3, m=0), 116 (r=4, m=0), etc. are

10 obtainable. Variations in additions to those set forth herein will be apparent to the skilled artisan.

Different mass-modified detector oligonucleotides can be used to simultaneously detect all possible variants/mutants simultaneously (FIGURE 6B). Alternatively, all four base permutations at the site of a

15 mutation can be detected by designing and positioning a detector oligonucleotide, so that it serves as a primer for a DNA/RNA polymerase with varying combinations of elongating and terminating nucleoside triphosphates (FIGURE 6C). For example, mass modifications also can be incorporated during the amplification process.

20 FIGURE 3 shows a different multiplex detection format, in which differentiation is accomplished by employing different specific capture sequences which are position-specifically immobilized on a flat surface (e.g., a 'chip array'). If different target sequences T1-Tn are present, their target capture sites TCS1-TCSn will specifically interact with

25 complementary immobilized capture sequences C1-Cn. Detection is achieved by employing appropriately mass differentiated detector oligonucleotides D1-Dn, which are mass modifying functionalities M1-Mn.

-66-

### Mass spectrometric methods for sequencing DNA

Amenable mass spectrometric formats for use herein include the ionization (I) techniques, such as matrix assisted laser desorption ionization (MALDI), electrospray (ESI) (e.g., continuous or pulsed); and 5 related methods (e.g., Ionspray, Thermospray, Fast Atomic Bombardment), and massive cluster impact (MCI); these ion sources can be matched with detection formats including lin-linear fields) time-of-flight (TOF), single or multiple quadrupole, single or multiple magnetic sector, Fourier transform ion cyclotron resonance (FTICR), ion trap, or 10 combinations of these to give a hybrid detector (e.g., ion trap - time of flight). For ionization, numerous matrix/wavelength combinations including frozen analyte preparation (MALDI) or solvent combinations (ESI) can be employed.

Since a normal DNA molecule includes four nucleotide units (A, T, 15 C, G), and the mass of each of these is unique (monoisotopic masses 313.06, 304.05, 289.05, 329.05 Da, respectively), an accurate mass determination can define or constrain the possible base compositions of that DNA. Only above 4900 Da does each unit molecular weight have at least one allowable composition; among all 5-mers there is only one non- 20 unique *nominal* molecular weight, among 8-mers, 20. For these and larger oligonucleotides, such mass overlaps can be resolved with the  $\sim 1/10^5$  ( $\sim 10$  part per million, ppm) mass accuracy available with high resolution FTICR MS. For the 25-mer  $A_5T_{20}$ , the 20 composition degeneracies when measured at  $\pm 0.5$  Da is reduced to three ( $A_5T_{20}$ , 25  $T_4C_{12}G_9$ ,  $AT_3C_4G_{16}$ ) when measured with 2 ppm accuracy. Given composition constraints (e.g., the presence or absence of one of the four bases in the strand) can reduce this further (see below).



-67-

Medium resolution instrumentation, including but not exclusively curved field reflectron or delayed extraction time-of-flight MS instruments, can also result in improved DNA detection for sequencing or diagnostics. Either of these are capable of detecting a 9 Da ( $\Delta m$  (A-T)) shift in  $\geq 30$ -mer strands generated from, for example primer oligo base extension (PROBE), or competitive oligonucleotide single base extension (COSBE), sequencing, or direct detection of small amplified products.

#### **BiomassScan**

In this embodiment, exemplified in Example 33, two single stranded nucleic acids are individually immobilized to solid supports. One support contains a nucleic acid encoding the wild type sequence whereas the other support contains a nucleic acid encoding a mutant target sequence. Total human genomic DNA is digested with one or more restriction endonuclease enzyme resulting in the production of small fragments of double stranded genomic DNA (10-1,000 bp). The digested DNA is incubated with the immobilized single stranded nucleic acids and the sample is heated to denature the DNA duplex. The immobilized nucleic acid competes with the other genomic DNA strand for the complementary DNA strand and under the appropriate conditions, a portion of the complementary DNA strand hybridizes to the immobilized nucleic acid resulting in a strand displacement. By using high stringency washing conditions, the two nucleic acids will remain as a DNA duplex only if there is exact identity between the immobilized nucleic acid and the genomic DNA strand. The DNA that remains hybridized to the immobilized nucleic acid is analyzed by mass spectrometry and detection of a signal in the mass spectrum of the appropriate mass is diagnostic for the wild type or mutant allele. In this manner, total genomic DNA can be isolated from a biological sample and screened for the presence or absence of certain mutations. By immobilizing a variety of single

-68-

stranded nucleic acids in an array format, a panel of mutations may be simultaneously screened for a number of genetic loci (i.e., multiplexing).

In addition, using less stringent washing conditions the hybridized DNA strand may be analyzed by mass spectrometry for changes in the  
5 mass resulting from a deletion or insertion within the targeted restriction endonuclease fragment.

#### Primer oligonucleotide base extension

As described in detail in the following Example 11, the primer oligo base extension (PROBE) method combined with mass spectrometry  
10 identifies the exact number of repeat units (i.e. the number of nucleotides in homogenous stretches) as well as second site mutations within a polymorphic region, which are otherwise only detectable by sequencing. Thus, the PROBE technique increases the total number of detectable alleles at a distinct genomic site, leading to a higher polymorphism  
15 information content (PIC) and yielding a far more definitive identification in for instance statistics-based analyses in paternity or forensics applications.

The method is based on the extension of a detection primer that anneals adjacent to a variable nucleotide tandem repeat (VNTR) or a  
20 polymorphic mononucleotide stretch using a DNA polymerase in the presence of a mixture of deoxyNTPs and those dideoxyNTPs that are not present in the deoxy form. The resulting products are evaluated and resolved by MALDI-TOF mass spectrometry without further labeling of the DNA. In a simulated routine application with 28 unrelated  
25 individuals, the mass error of this procedure using external calibration was in the worst case 0.38% (56-mer), which is comparable to approximately 0.1 base accuracy; routine standard mass deviations are in the range of 0.1% (.03 bases). Such accuracy with conventional

-69-

electrophoretic methods is not realistic, underscoring the value of PROBE and mass spectrometry in forensic medicine and paternity testing.

The ultra-high resolution of Fourier Transform mass spectrometry makes possible the simultaneous measurement of all reactions of a  
5 Sanger or Maxam Gilbert sequencing experiment, since the sequence may be read from mass differences instead of base counting from 4 tubes.

Additionally, the mass differences between adjacent bases generated from unilateral degradation in a stepwise manner by an  
10 exonuclease can be used to read the entire sequence of fragments generated. Whereas UV or fluorescent measurements will not discriminate mixtures of the nucleoside/nucleotide which are generated when the exonuclease enzyme gets out of phase, this is no problem with mass spectrometry since the resolving power in differentiating between  
15 the molecular mass of dA, dT, dG and dC is more than significant. The mass of the adjacent bases (i.e., nucleotides) can be determined, for example, using Fast Atomic Bombardment (FAB) or Electrospray Ionization (ESI) mass spectrometry.

New mutation screening over an entire amplified product can be  
20 achieved by searching for mass shifted fragments generated in an endonuclease digestion as described in detail in the following Examples 4 and 12.

Partial sequence information obtained from tandem mass spectrometry ( $MS^n$ ) can place composition constraints as described in the  
25 preceding paragraph. For the 25-mer above, generation of two fragment ions formed by collisionally activated dissociation (CAD) which differ by 313 Da discards  $T_4C_{12}G_9$ , which contains no A nucleotides; confirming more than a single A eliminates  $AT_3C_4G_{16}$  as a possible composition.

-70-

MS<sup>n</sup> can also be used to determine full or partial sequences of larger DNAs; this can be used to detect, locate, and identify new mutations in a given gene region. Enzymatic digest products whose masses are correct need not be further analyzed; those with mass shifts could be isolated in real time from the complex mixture in the mass spectrometer and partially sequenced to locate the new mutation.

Table I describes the mutation/polymorphism detection tests that have been developed.

**Table I**  
**Mutation/Polymorphism Detection Tests**

Clinical Association	Gene	Mutation/Polymorphism
Cystic Fibrosis	CFTR	38 disease causing mutations in 14 exons/introns
Heart Disease (Cholesterol Metabolism)	Apo E Apo A-IV Apo B-100	112R, 112C, 158R, 158C 347S, 347T, 360H, 360Q 3500Q, 3500R
Thyroid Cancer	RET proto-oncogene	C634W, C634T, C634R, C634S, C634F
Sickle Cell Anemia/Thalassemia	beta-globin	Sickle cell anemia S and C 45 thalassemia alleles
HIV Susceptibility	CCR-5	32bp deletion
Breast Cancer Susceptibility	BRCA-2	2bp (AG) deletion in exon 2
Thrombosis	Factor V	R506Q
Arteriosclerosis	GpIIb E-selectin	L33P S128R
Hypertension	ACE	I/D polymorphism

## 25 Detection of mutations

### Diagnosis of genetic diseases

The mass spectrometric processes described above can be used, for example, to diagnose any of the more than 3000 genetic diseases

-71-

currently known (e.g., hemophilias, thalassemias, Duchenne Muscular Dystrophy (DMD), Huntington's Disease (HD), Alzheimer's Disease and Cystic Fibrosis (CF)) or to be identified.

The following Example 3 provides a mass spectrometric method for detecting a mutation ( $\Delta F508$ ) of the cystic fibrosis transmembrane conductance regulator gene (CFTR), which differs by only three base pairs (900 daltons) from the wild type of CFTR gene. As described further in Example 3, the detection is based on a single-tube, competitive oligonucleotide single base extension (COSBE) reaction using a pair of primers with the 3'-terminal base complementary to either the normal or mutant allele. Upon hybridization and addition of a polymerase and the nucleoside triphosphate one base downstream, only those primers properly annealed (i.e., no 3'-terminal mismatch) are extended; products are resolved by molecular weight shifts as determined by matrix assisted laser desorption ionization time-of-flight mass spectrometry. For the cystic fibrosis  $\Delta F508$  polymorphism, 28-mer 'normal' (N) and 30-mer 'mutant' (M) primers generate 29- and 31-mers for N and M homozygotes, respectively, and both for heterozygotes. Since primer and product molecular weights are relatively low (<10 kDa) and the mass difference between these are at least that of a single ~300 Da nucleotide unit, low resolution instrumentation is suitable for such measurements.

Thermosequence cycle sequencing, as further described in Example 11, is also useful for detecting a genetic disease.

In addition to mutated genes, which result in genetic disease, certain birth defects are the result of chromosomal abnormalities such as Trisomy 21 (Down's Syndrome), Trisomy 13 (Patau Syndrome), Trisomy 18 (Edward's Syndrome), Monosomy X (Turner's Syndrome) and other sex chromosome aneuploidies such as Klienfelter's Syndrome (XXY).

-72-

Here, "house-keeping" genes encoded by the chromosome in question are present in different quantity and the different amount of an amplified fragment compared to the amount in a normal chromosomal configuration can be determined by mass spectrometry.

- 5 Further, there is growing evidence that certain DNA sequences may predispose an individual to any of a number of diseases such as diabetes, arteriosclerosis, obesity, various autoimmune diseases and cancer (e.g., colorectal, breast, ovarian, lung). Also, the detection of "DNA fingerprints", e.g., polymorphisms, such as "mini- and micro-
- 10 satellite sequences", are useful for determining identity or heredity (e.g., paternity or maternity).

The following Examples 4 and 12 provide mass spectrometer based methods for identifying any of the three different isoforms of human apolipoprotein E, which are coded by the E2, E3 and E4 alleles.

- 15 For example, the molecular weights of DNA fragments obtained after restriction with appropriate restriction endonucleases can be used to detect the presence of a mutation and/or a specific allele.

Depending on the biological sample, the diagnosis for a genetic disease, chromosomal aneuploidy or genetic predisposition can be

20 preformed either pre- or post-natally.

#### Diagnosis of cancer

- Preferred mass spectrometer-based methods for providing an early indication of the existence of a tumor or a cancer are provide herein. For example, as described in Example 13, the telomeric repeat amplification
- 25 protocol (TRAP) in conjunction with telomerase specific extension of a substrate primer and a subsequent amplification of the telomerase specific extension products by an amplification step using a second primer complementary to the repeat structure was used to obtain extension ladders, that were easily detected by MALDI-TOF mass

-73-

spectrometry as an indication of telomerase activity and therefor tumorigenesis.

Alternatively, as described in Example 14, expression of a tumor or cancer associated gene (e.g., human tyrosine 5-hydroxylase) via RT-PCR  
5 and analysis of the amplified products by mass spectrometry can be used to detect the tumor or cancer (e.g., biosynthesis of catecholamine via tyrosine 5-hydroxylase is a characteristic of neuroblastoma).

Further, a primer oligo base extension reaction and detection of products by mass spectrometry provides a rapid means for detecting the  
10 presence of oncogenes, such as the RET proto oncogene codon 634, which is related to causing multiple endocrine neoplasia, type II (MEN II), as described in Example 15.

#### Diagnosis of infection

Viruses, bacteria, fungi and other infectious organisms contain  
15 distinct nucleic acid sequences, which are different from the sequences contained in the host cell. Detecting or quantitating nucleic acid sequences that are specific to the infectious organism is important for diagnosing or monitoring infection. Examples of disease causing viruses that infect humans and animals and which may be detected by the  
20 disclosed processes include: *Retroviridae* (e.g., human immunodeficiency viruses, such as HIV-1 (also referred to as HTLV-III, LAV or HTLV-III/LAV, see, e.g., Ratner et al. (1985) Nature 313: 227-284; Wain-Hobson et al. (1985) Cell 40:9-17); HIV-2 (see, Guyader et al. (1987) Nature 328:662-669 European Patent Publication No. 0 269 520;  
25 Chakrabarti et al. (1987) Nature 328:543-547; and European Patent Application No. 0 655 501); and other isolates, such as HIV-LP (International PCT application No. WO 94/00562 entitled "A Novel Human Immunodeficiency Virus"; *Picornaviridae* (e.g., polio viruses, hepatitis A virus, (se , e.g., Gust et al. (1983) Intervirology 20:1-7);

- entero viruses, human coxsackie viruses, rhinoviruses, echoviruses); *Calciviridae* (e.g., strains that cause gastroenteritis); *Togaviridae* (e.g., equine encephalitis viruses, rubella viruses); *Flaviridae* (e.g., dengue viruses, encephalitis viruses, yellow fever viruses); *Coronaviridae* (e.g., 5 coronaviruses); *Rhabdoviridae* (e.g., vesicular stomatitis viruses, rabies viruses); *Filoviridae* (e.g., ebola viruses); *Paramyxoviridae* (e.g., parainfluenza viruses, mumps virus, measles virus, respiratory syncytial virus); *Orthomyxoviridae* (e.g., influenza viruses); *Bunyaviridae* (e.g., Hantaan viruses, bunya viruses, phleboviruses and Nairo viruses); *Arena* 10 *viridae* (hemorrhagic fever viruses); *Reoviridae* (e.g., reoviruses, orbiviruses and rotaviruses); *Birnaviridae*; *Hepadnaviridae* (Hepatitis B virus); *Parvoviridae* (parvoviruses); *Papovaviridae* (papilloma viruses, polyoma viruses); *Adenoviridae* (most adenoviruses); *Herpesviridae* (herpes simplex virus (HSV) 1 and 2, varicella zoster virus, 15 cytomegaovirus (CMV), herpes viruses'); *Poxviridae* (variola viruses, vaccinia viruses, pox viruses); and *Iridoviridae* (e.g., African swine fever virus); and unclassified viruses (e.g., the etiological agents of Spongiform encephalopathies, the agent of delta hepatitis (thought to be a defective satellite of hepatitis B virus), the agents of non-A, non-B hepatitis (class 20 1 = internally transmitted; class 2 = parenterally transmitted (i.e., Hepatitis C); Norwalk and related viruses, and astroviruses).

Examples of infectious bacteria include, but are not limited to:

- Helicobacter pylori*, *Borrelia burgdorferi*, *Legionella pneumophila*, *Mycobacteria* sps (e.g., *M. tuberculosis*, *M. avium*, *M. intracellulare*, *M.* 25 *kansasii*, *M. gordonae*), *Staphylococcus aureus*, *Neisseria gonorrhoeae*, *Neisseria meningitidis*, *Listeria monocytogenes*, *Streptococcus pyogenes* (Group A Streptococcus), *Streptococcus agalactiae* (Group B Streptococcus), *Streptococcus* (viridans group), *Streptococcus faecalis*, *Streptococcus bovis*, *Streptococcus* (anaerobic sps.), *Streptococcus*



-75-

*phhenumoniae*, pathogenic *Campylobacter* sp., *Enterococcus* sp.,  
*Haemophilus influenzae*, *Bacillus anthracis*, *corynebacterium diphtheriae*,  
*corynebacterium* sp., *Erysipelothrix rhusiopathiae*, *Clostridium*  
*perfringers*, *Clostridium tetani*, *Enterobacter aerogenes*, *Klebsiella*  
5 *pneumoniae*, *Pasturella multocida*, *Bacteroides* sp., *Fusobacterium*  
*nucleatum*, *Streptobacillus moniliformis*, *Treponema pallidum*,  
*Treponema pertenuae*, *Leptospira*, and *Actinomyces israeli*.

Examples of infectious fungi include: *Cryptococcus neoformans*,  
*Histoplasma capsulatum*, *Coccidioides immitis*, *Blastomyces dermatitidis*,  
10 *Chlamydia trachomatis*, *Candida albicans*. Other infectious organisms  
(i.e., protists) include: *Plasmodium falciparum* and *Toxoplasma gondii*.

The processes provided herein makes use of the known sequence  
information of the target sequence and known mutation sites. Although  
new mutations can also be detected. For example, as shown in FIGURE  
15 8, transcription of a nucleic acid molecule obtained from a biological  
sample can be specifically digested using one or more nucleases and the  
fragments captured on a solid support carrying the corresponding  
complementary nucleic acid sequences. Detection of hybridization and  
the molecular weights of the captured target sequences provide  
20 information on whether and where in a gene a mutation is present.  
Alternatively, DNA can be cleaved by one or more specific  
endonucleases to form a mixture of fragments. Comparison of the  
molecular weights between wildtype and mutant fragment mixtures  
results in mutation detection.

#### 25 Sequencing by generation of specifically terminated fragments

In another embodiment, an accurate sequence determination of a  
relatively large target nucleic acid, can be obtained by generating  
specifically terminated fragments from the target nucleic acid,  
determining the mass of each fragment by mass spectrometry and

-76-

ordering the fragments to determine the sequence of the larger target nucleic acid. In a preferred embodiment, the specifically terminated fragments are partial or complete base-specifically terminated fragments.

- One method for generating base specifically terminated fragments involves using a base-specific ribonuclease after e.g., a transcription reaction. Preferred base-specific ribonucleases are selected from among: T<sub>1</sub>-ribonuclease (G-specific), U<sub>2</sub>-ribonuclease (A-specific), PhyM-ribonuclease U specific and ribonuclease A (U/C specific). Other efficient and base-specific ribonucleases can be identified using the assay described in Example 16. Preferably modified nucleotides are included in the transcription reaction with unmodified nucleotides. Most preferably, the modified nucleotides and unmodified nucleotides are added to the transcription reaction at appropriate concentrations, so that both moieties are incorporated at a preferential rate of about 1:1. Alternatively, two separate transcriptions of the target DNA sequence one with the modified and one with the unmodified nucleotides can be performed and the results compared. Preferred modified nucleotides include: boron or bromine modified nucleotides (Porter et al. (1995) Biochemistry 34:11963-11969; Hasan et al. (1996) Nucl. Acids Res. 24:2150-2157; Li et al. (1995) Nucleic Acids Res. 23:4495-4501),  $\alpha$ -thio-modified nucleotides, as well as mass-modified nucleotides as described above.

- Another method for generating base specifically terminated fragments involves performing a combined amplification and base-specific termination reaction. For example, a combined amplification and termination reaction can be performed using at least two different polymerase enzymes, each having a different affinity for the chain terminating nucleotide, so that polymerization by an enzyme with relatively low affinity for the chain terminating nucleotide leads to exponential amplification whereas an enzyme with relatively high affinity

-77-

for the chain terminating nucleotide terminates the polymerization and yields sequencing products.

- The combined amplification and sequencing can be based on any amplification procedure that employs an enzyme with polynucleotide synthetic ability (e.g., polymerase). One preferred process, based on the polymerase chain reaction (PCR), includes the following three thermal steps: 1) denaturing a double stranded (ds) DNA molecule at an appropriate temperature and for an appropriate period of time to obtain the two single stranded (ss) DNA molecules (the template: sense and antisense strand); 2) contacting the template with at least one primer that hybridizes to at least one ss DNA template at an appropriate temperature and for an appropriate period of time to obtain a primer containing ss DNA template; 3) contacting the primer containing template at an appropriate temperature and for an appropriate period of time with:
- (i) a complete set of chain elongating nucleotides, (ii) at least one chain terminating nucleotide, (iii) a first DNA polymerase, which has a relatively low affinity towards the chain terminating nucleotide; and (iv) a second DNA polymerase, which has a relatively high affinity towards the chain terminating nucleotide.
- Steps 1)-3) can be sequentially performed for an appropriate number of times (cycles) to obtain the desired amount of amplified sequencing ladders. The quantity of the base specifically terminated fragment desired dictates how many cycles are performed. Although an increased number of cycles results in an increased level of amplification, it may also detract from the sensitivity of a subsequent detection. It is therefore generally undesirable to perform more than about 50 cycles, and is more preferable to perform less than about 40 cycles (e.g., about 20-30 cycles).

-78-

Another preferred process for simultaneously amplifying and chain terminating a nucleic acid sequence is based on strand displacement amplification (SDA) (see, e.g., Walker *et al.* (1994) *Nucl. Acids Res.* 22:2670-77; European Patent Publication Number 0 684 315 entitled "Strand Displacement Amplification Using Thermophilic Enzymes"). In essence, this process involves the following three steps, which altogether constitute a cycle: 1) denaturing a double stranded (ds) DNA molecule containing the sequence to be amplified at an appropriate temperature and for an appropriate period of time to obtain the two single stranded (ss) DNA molecules (the template: sense and antisense strand); 2) contacting the template with at least one primer (P), that contains a recognition/cleavage site for a restriction endonuclease (RE) and that hybridizes to at least one ss DNA template at an appropriate temperature and for an appropriate period of time to obtain a primer containing ss DNA template; 3) contacting the primer containing template at an appropriate temperature and for an appropriate period of time with (i) a complete set of chain elongating nucleotides; (ii) at least one chain terminating nucleotide; (iii) a first DNA polymerase, which has a relatively low affinity towards the chain terminating nucleotide; (iv) a second DNA polymerase, which has a relatively high affinity towards the chain terminating nucleotide; and (v) an RE that nicks the primer recognition/cleavage site.

Steps 1)-3) can be sequentially performed for an appropriate number of times (cycles) to obtain the desired amount of amplified sequencing ladders. As with the PCR based process, the quantity of the base specifically terminated fragment desired dictates how many cycles are performed. Preferably, less than 50 cycles, more preferably less than about 40 cycles and most preferably about 20 to 30 cycles are performed.

-79-

Preferably about 0.5 to about 3 units of polymerase is used in the combined amplification and chain termination reaction. Most preferably about 1 to 2 units is used. Particularly preferred polymerases for use in conjunction with PCR or other thermal amplification process are

- 5 thermostable polymerases, such as Taq DNA polymerase (Boehringer Mannheim), AmpliTaq FS DNA polymerase (Perkin-Elmer), Deep Vent (exo-), Vent, Vent (exo-) and Deep Vent DNA polymerases (New England Biolabs), Thermo Sequenase (Amersham) or *exo(-) Pseudococcus furiosus (Pfu)* DNA polymerase (Stratagene, Heidelberg, Germany).
- 10 AmpliTaq, Ultman, 9 degree Nm, Tth, Hot Tub, and *Pyrococcus furiosus*. In addition, preferably the polymerase does not have 5'-3' exonuclease activity.

- In addition to polymerases, which have a relatively high and a relatively low affinity to the chain terminating nucleotide, a third
- 15 polymerase, which has proofreading capacity (e.g., *Pyrococcus woesei (Pwo)*) DNA polymerase may also be added to the amplification mixture to enhance the fidelity of amplification.

- Yet another method for generating base specifically terminated fragments involves contacting an appropriate amount of the target
- 20 nucleic acid with a specific endonuclease or exonuclease. Preferably, the original 5' and/or 3' end of the nucleic acid is tagged to facilitate the ordering of fragments. Tagging of the 3' end is particularly preferred when *in vitro* nucleic acid transcripts are being analyzed, so that the influence of 3' heterogeneity, premature termination and nonspecific
- 25 elongation can be minimized. 5' and 3' tags can be natural (e.g., a 3' poly A tail or 5' or 3' heterogeneity) or artificial. Preferred 5' and/or 3' tags are selected from among the molecules described for mass-modification above.

The methods provided herein are further illustrated by the following examples, which should not be construed as limiting in any way.

#### EXAMPLE 1

##### 5 MALDI-TOF desorption of oligonucleotides directly on solid supports

1 g CPG (Controlled Pore Glass) was functionalized with 3-(triethoxysilyl)-epoxypropan to form OH-groups on the polymer surface. A standard oligonucleotide synthesis with 13 mg of the OH-CPG on a DNA synthesizer (Milligen, Model 7500) employing  $\beta$ -cyanoethyl-phosphoamidites (Köster *et al.* (1994) Nucleic Acids Res. 12:4539) and TAC N-protecting groups (Köster *et al.* (1981) Tetrahedron 37:362) was performed to synthesize a 3'-T<sub>6</sub>-50mer oligonucleotide sequence in which 50 nucleotides are complementary to a "hypothetical" 50mer sequence. T<sub>6</sub> serves as a spacer. Deprotection with saturated ammonia in methanol at room temperature for 2 hours furnished according to the determination of the DMT group CPG which contained about 10  $\mu$ mol 55mer/g CPG. This 55mer served as a template for hybridizations with a 26-mer (with 5'-DMT group) and a 40-mer (without DMT group). The reaction volume is 100  $\mu$ l and contains about 1 nmol CPG bound 55mer as template, an equimolar amount of oligonucleotide in solution (26-mer or 40-mer) in 20 mM Tris-HCl, pH 7.5, 10 mM MgCl<sub>2</sub> and 25 mM NaCl. The mixture was heated for 10 min at 65°C and cooled to 37°C during 30' (annealing). The oligonucleotide which has not been hybridized to the polymer-bound template were removed by centrifugation and three subsequent washing/centrifugation steps with 100  $\mu$ l each of ice-cold 50 mM ammoniumcitrate. The beads were air-dried and mixed with matrix solution (3-hydroxypicolinic acid/10mM ammonium citrate in acetonitrile/water, 1:1), and analyzed by MALDI-TOF mass spectrometry. The results are presented in Figures 10 and 11.

-81-

**EXAMPLE 2**

**Electrospray (ES) desorption and differentiation of an 18-mer and 19-mer DNA fragments at a concentration of 50 pmole/ $\mu$ l in**

**2-propanol/10mM ammoniumcarbonate (1/9, v/v) were analyzed**

**5 simultaneously by an electrospray mass spectrometer.**

**The successful desorption and differentiation of an 18-mer and 19-mer by electrospray mass spectrometry is shown in FIGURE 12.**

**EXAMPLE 3**

**10 Detection of The Cystic Fibrosis Mutation  $\Delta$ F508, by single step dideoxy extension and analysis by MALDI-TOF mass spectrometry (Competitive Oligonucleotide Simple Base Extension = COSBE)**

**The principle of the COSBE method is shown in FIGURE 13, N being the normal and M the mutation detection primer, respectively.**

**MATERIALS AND METHODS**

**15       *PCR Amplification and Strand Immobilization.*** Amplification was carried out with exon 10 specific primers using standard PCR conditions (30 cycles: 1'@95°C, 1'@55°C, 2'@72°C); the reverse primer was 5' labelled with biotin and column purified (Oligopurification Cartridge, Cruachem). After amplification the amplified products were purified by  
**20 column separation (Qiagen Quickspin) and immobilized on streptavidin coated magnetic beads (Dynabeads, Dynal, Norway) according to their standard protocol; DNA was denatured using 0.1 M NaOH and washed with 0.1M NaOH, 1xB+W buffer and TE buffer to remove the non-biotinylated sense strand.**

**25       *COSBE Conditions.*** The beads containing ligated antisense strand were resuspended in 18 $\mu$ l of Reaction mix 1 (2  $\mu$ l 10X Taq buffer, 1  $\mu$ L (1 unit) Taq Polymerase, 2  $\mu$ L of 2 mM dGTP, and 13  $\mu$ L H<sub>2</sub>O) and incubated at 80°C for 5' before the addition of Reaction mix 2 (100 ng each of COSBE primers). The temperature was reduced to 60°C and the  
**30 mixtures incubated for a 5' annealing/extension period; the beads were**

then washed in 25mM triethylammonium acetate (TEAA) followed by 50mM ammonium citrate.

*Primer Sequences.* All primers were synthesized on a Perseptive Biosystems Expedite 8900 DNA Synthesizer using conventional

5 phosphoramidite chemistry (Sinha *et al.* (1984) Nucleic Acids Res. 12:4539). COSBE primers (each containing an intentional mismatch one base before the 3'-terminus) were those used in a previous ARMS study (Ferrie *et al.* (1992) Am J Hum Genet 51:251-262) with the exception that two bases were removed from the 5'-end of the normal:

10 Ex10 PCR (Forward): 5'-BIO-GCA AGT GAA TCC TGA GCG TG-3'  
(SEQ ID No. 1)

Ex10 PCR (Reverse): 5'-GTG TGA AGG GTT CAT ATG C-3'  
(SEQ ID No. 2)

COSBE ΔF508-N 5'-ATC TAT ATT CAT CAT AGG AAA CAC CAC A-3'  
15 (28-mer) (SEQ ID No. 3)

COSBE ΔF508-N 5'-GTA TCT ATA TTC ATC ATA GGA AAC ACC ATT-  
3' (30-mer) (SEQ ID No. 4)

*Mass Spectrometry.* After washing, beads were resuspended in 1 μL 18 Mohm/cm H<sub>2</sub>O. 300 nL each of matrix (Wu *et al.* (1993) Rapid Commun. Mass Spectrom. 7:142-146) solution (0.7 M 3-hydroxypicolinic acid, 0.7 M dibasic ammonium citrate in 1:1 H<sub>2</sub>O:CH<sub>3</sub>CN) and resuspended beads (Tang *et al.* (1995) Rapid Commun Mass Spectrom 8:727-730) were mixed on a sample target and allowed to air dry. Up to 20 samples were spotted on a probe target disk for introduction into the source region of an unmodified Thermo Bioanalysis (formerly Finnigan) Visions 2000 MALDI-TOF operated in reflectron mode with 5 and 20 kV on the target and conversion dynode, respectively. Theoretical average molecular weights (M<sub>r</sub>(calc)) were calculated from atomic compositions. Vendor provided software was used to determine peak center IDs using



-83-

external calibration; 1.08 Da has been subtracted from these to correct for the charge carrying proton mass to yield the text  $M_{(exp)}$  values.

**Scheme.** Upon annealing to the bound template, the N and M primers (8508.6 and 9148.0 Da, respectively) are presented with dGTP;

- 5 only primers with proper Watson-Crick base pairing at the variable (V) position are extended by the polymerase. Thus if V pairs with the 3'-terminal base of N, N is extended to a 8837.9 Da product (N + 1). Likewise, if V is properly matched to the M terminus, M is extended to a 9477.3 Da M + 1 product.

10 **Results**

Figures 14 - 18 show the representative mass spectra of COSBE reaction products. Better results were obtained when amplified products were purified before the biotinylated anti-sense strand was bound.

**EXAMPLE 4**

15 **Differentiation of Human Apolipoprotein E Isoforms by Mass Spectrometry**

Apolipoprotein E (Apo E), a protein component of lipoproteins, plays an essential role in lipid metabolism. For example, it is involved with cholesterol transport, metabolism of lipoprotein particles,

- 20 immunoregulation and activation of a number of lipolytic enzymes.

There are three common isoforms of human Apo E (coded by E2, E3 and E4 alleles). The most common is the E3 allele. The E2 allele has been shown to decrease the cholesterol level in plasma and therefore may have a protective effect against the development of atherosclerosis.

- 25 The DNA encoding a portion of the E2 allele is set forth in SEQ ID No. 130. Finally, the E4 isoform has been correlated with increased levels of cholesterol, conferring predisposition to atherosclerosis. Therefore, the identity of the apo E allele of a particular individual is an important determinant of risk for the development of cardiovascular disease.

-84-

As shown in Figure 19, a sample of DNA encoding apolipoprotein E can be obtained from a subject, amplified (e.g., via PCR); and the amplified product can be digested using an appropriate enzyme (e.g., CfoI). The restriction digest obtained can then be analyzed by a variety of means. As shown in Figure 20, the three isotypes of apolipoprotein E (E2, E3 and E4 have different nucleic acid sequences and therefore also have distinguishable molecular weight values.

As shown in Figure 21A-C, different Apolipoprotein E genotypes exhibit different restriction patterns in a 3.5% MetPhor Agarose Gel or 12% polyacrylamide gel. As shown in Figures 22 and 23, the various apolipoprotein E genotypes can also be accurately and rapidly determined by mass spectrometry.

#### EXAMPLE 5

Detection of hepatitis B virus in serum samples.

#### 15 MATERIALS AND METHODS

##### *Sample preparation*

Phenol/choloform extraction of viral DNA and the final ethanol precipitation was done according to standard protocols.

##### *First PCR*

Each reaction was performed with 5 $\mu$ l of the DNA preparation from serum. 15 pmol of each primer and 2 units Taq DNA polymerase (Perkin Elmer, Weiterstadt, Germany) were used. The final concentration of each dNTP was 200 $\mu$ MM, the final volume of the reaction was 50  $\mu$ l. 10x PCR buffer (Perkin Elmer, Weiterstadt, Germany) contained 100 mM Tris-HCl, pH 8.3, 500 mM KCl, 15 mM MgCl<sub>2</sub>, 0.01% gelatine (w/v).

Primer sequences:

Primer	SEQUENCE	SEQ ID No.
1	5'-GCTTTGGGGCATGGACATTGACCCGTATAA-3'	5
2	5'-CTGACTACTAATTCCTGGATGCTGGGTCT-3'	6

-85-

*Nested PCR:*

Each reaction was performed either with 1  $\mu$ l of the first reaction or with a 1:10 dilution of the first PCR as template, respectively. 100 pmol of each primer, 2.5 u *Pfu*(exo-) DNA polymerase (Stratagene, Heidelberg, Germany), a final concentration of 200  $\mu$ M of each dNTPs and 5  $\mu$ l 10x *Pfu* buffer (200 mM Tris-HCl, pH 8.75, 100 mM KCl, 100 mM (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, 1% Triton X-100, 1 mg/ml BSA, (Stratagene, Heidelberg, Germany) were used in a final volume 50  $\mu$ l. The reactions were performed in a thermocycler (OmniGene, MWG-Biotech, Ebersberg, Germany) using the following program: 92°C for 1 minute, 60°C for 1 minute and 72°C for 1 minute with 20 cycles. Sequence of oligodeoxynucleotides (purchased HPLC-purified from MWG-Biotech, Ebersberg, Germany):

- HBV13: 5'-TTGCCTGAGTGCAGTATGGT-3' (SEQ ID NO. 7)  
15 HBV15bio: Biotin-5'-AGCTCTATATCGGGAAGCCT-3' (SEQ ID NO. 8)

*Purification of amplified products:*

- For the recording of each spectrum, one PCR, 50  $\mu$ l, (performed as described above) was used. Purification was done according to the following procedure: Ultrafiltration was done using Ultrafree-MC filtration units (Millipore, Eschborn, Germany) according to the protocol of the provider with centrifugation at 8000 rpm for 20 minutes. 25 $\mu$ l (10 $\mu$ g/ $\mu$ l) streptavidin Dynabeads (Dyna, Hamburg, Germany) were prepared according to the instructions of the manufacturer and resuspended in 25 $\mu$ l of B/W buffer (10 mM Tris-HCl, pH 7.5, 1mM EDTA, 2 M NaCl).  
25 This suspension was added to the PCR samples still in the filtration unit and the mixture was incubated with gentle shaking for 15 minutes at ambient temperature. The suspension was transferred in a 1.5 ml Eppendorf tube and the supernatant was removed with the aid of a Magnetic Particle Collector, MPC, (Dyna, Hamburg, Germany). The

-86-

beads were washed twice with 50  $\mu$ l of 0.7 M ammonium citrate solution, pH 8.0 (the supernatant was removed each time using the MPC). Cleavage from the beads can be accomplished by using formamide at 90°C. The supernatant was dried in a speedvac for about 5 an hour and resuspended in 4  $\mu$ l of ultrapure water (MilliQ UF plus Millipore, Eschborn, Germany). This preparation was used for MALDI-TOF MS analysis.

#### *MALDI-TOF MS:*

Half a microliter of the sample was pipetted onto the sample 10 holder, then immediately mixed with 0.5  $\mu$ l matrix solution (0.7 M3-hydroxypicolinic acid 50% acetonitrile, 70 mM ammonium citrate). This mixture was dried at ambient temperature and introduced into the mass spectrometer. All spectra were taken in positive ion mode using a Finnigan MAT Vision 2000 (Finnigan MAT, Bremen, Germany), equipped 15 with a reflectron (5 keV ion source, 20 keV postacceleration) and a 337 nm nitrogen laser. Calibration was done with a mixture of a 40-mer and a 100-mer. Each sample was measured with different laser energies. In the negative samples, the amplified product was detected neither with less nor with higher laser energies. In the positive samples the amplified 20 product was detected at different places of the sample spot and also with varying laser energies.

#### **RESULTS**

A nested PCR system was used for the detection of HBV DNA in blood samples employing oligonucleotides complementary to the c region 25 of the HBV genome (primer 1: beginning at map position 1763, primer 2 beginning at map position 2032 of the complementary strand) encoding the HBV core antigen (HBVcAg). DNA was isolated from patients serum according to standard protocols. A first PCR was performed with the

-87-

DNA from these preparations using a first set of primers. If HBV DNA was present in the sample a DNA fragment of 269 bp was generated.

In the second reaction, primers which were complementary to a region within the PCR fragment generated in the first PCR were used. If  
5 HBV related amplified products were present in the first PCR a DNA fragment of 67 bp was generated (see Fig. 25A) in this nested PCR. The usage of a nested PCR system for detection provides a high sensitivity and also serves as a specificity control for the external PCR (Rolfs et al. (1992) PCR: Clinical Diagnostics and Research, Springer, Heidelberg). A  
10 further advantage is that the amount of fragments generated in the second PCR is high enough to ensure an unproblematic detection although purification losses can not be avoided.

The samples were purified using ultrafiltration to restreptavidin Dynabeads. This purification was done because the shorter primer  
15 fragments were immobilized in higher yield on the beads due to steric reasons. The immobilization was done directly on the ultrafiltration membrane to avoid substance losses due to unspecific absorption on the membrane. Following immobilization, the beads were washed with ammonium citrate to perform cation exchange (Pieles et al. (1993) Nucl.  
20 Acids Res. 21:3191-3196). The immobilized DNA was cleaved from the beads using 25% ammonia which allows cleavage of DNA from the beads in a very short time, but does not result in an introduction of sodium or other cations.

The nested PCRs and the MALDI TOF analysis were performed  
25 without knowing the results of serological analysis. Due to the unknown virus titer, each sample of the first PCR was used undiluted as template and in a 1:10 dilution, respectively.

Sample 1 was collected from a patient with chronic active HBV infection who was positive in Hbs- and Hbe-antigen tests but negative in

-88-

a dot blot analysis. Sample 2 was a serum sample from a patient with an active HBV infection and a massive viremia who was HBV positive in a dot blot analysis. Sample 3 was a denatured serum sample therefore no serological analysis could be performed by an increased level of transaminases indicating liver disease was detected. In autoradiograph analysis (Figure 24), the first PCR of this sample was negative. Nevertheless, there was some evidence of HBV infection. This sample is of interest for MALDI-TOF analysis, because it demonstrates that even low-level amounts of amplified products can be detected after the purification procedure. Sample 4 was from a patient who was cured of HBV infection. Samples 5 and 6 were collected from patients with a chronic active HBV infection.

Figure 24 shows the results of a PAGE analysis of the nested PCR reaction. A amplified product is clearly revealed in samples 1, 2, 3, 5 and 6. In sample 4 no amplified product was generated, it is indeed HBV negative, according to the serological analysis. Negative and positive controls are indicated by + and -, respectively. Amplification artifacts are visible in lanes 2, 5, 6 and + if non-diluted template was used. These artifacts were not generated if the template was used in a 1:10 dilution. In sample 3, amplified product was merely detectable if the template was not diluted. The results of PAGE analysis are in agreement with the data obtained by serological analysis except for sample 3 as discussed above.

Figure 25A shows a mass spectrum of a nested amplified product from sample number 1 generated and purified as described above. The signal at 20754 Da represents the single stranded amplified product (calculated: 20735 Da, as the average mass of both strands of the amplified product cleaved from the beads). The mass difference of calculated and obtained mass is 19 Da (0.09%). As shown in Fig. 25A,

-89-

sample number 1 generated a high amount of amplified product, resulting in an unambiguous detection.

Fig. 25B shows a spectrum obtained from sample number 3. As depicted in Fig. 24, the amount of amplified product generated in this section is significantly lower than that from sample number 1.

Nevertheless, the amplified product is clearly revealed with a mass of 20751 Da (calculated 20735). The mass difference is 16 Da (0.08%). The spectrum depicted in Fig. 25C was obtained from sample number 4 which is HBV negative (as is also shown in Fig 24). As expected no signals corresponding to the amplified product could be detected. All samples shown in Fig. 25 were analyzed with MALDI-TOF MS, whereby amplified product was detected in all HBV positive samples, but not in the HBV negative samples. These results were reproduced in several independent experiments.

15

## EXAMPLE 6

### Analysis of Ligase Chain Reaction Products Via MALDI-TOF Mass Spectrometry

#### MATERIALS AND METHODS

##### *Oligodeoxynucleotides*

20

Except the biotinylated one and all other oligonucleotides were synthesized in a 0.2  $\mu$ mol scale on a MilliGen 7500 DNA Synthesizer (Millipore, Bedford, MA, USA) using the  $\beta$ -cyanoethylphosphoramidite method (Sinha, N.D. *et al.* (1984) *Nucleic Acids Res.* 12:4539-4577).

The oligodeoxynucleotides were RP-HPLC-purified and deprotected according to standard protocols. The biotinylated oligodeoxynucleotide was purchased (HPLC-purified) from Biometra, Gottingen, Germany).

25

Sequences and calculated masses of the oligonucleotides used:

Olig deoxy- nucle tide	SEQUENCE	SEQ ID No.
30 A	5'-p-TTGTGCCACGCGTTGGGAATGTA (7521 Da)	9

-90-

B	5'-p-AGCAACGACTGTTTGCCCGCCAGTTG (7948 Da)	10
C	5'-bio-TACATTCCCAACCGCTGGCACAAC (7960 Da)	11
D	5'-p-AACTGGCGGGCAAACAGTCGTTGCT (7708 Da)	12

### 5 *5-Phosphorylation of oligonucleotides A and D*

This was performed with polynucleotide kinase (Boehringer, Mannheim, Germany) according to published procedures, the 5'-phosphorylated oligonucleotides were used unpurified for LCR.

### *Ligase chain reaction*

10 The LCR was performed with *Pfu* DNA ligase and a ligase chain reaction kit (Stratagene, Heidelberg, Germany) containing two different pBluescript KII phagemids. One carrying the wildtype form of the *E.coli lacI* gene and the other one a mutant of this gene with a single point mutation at bp 191 of the *lacI* gene.

15 The following LCR conditions were used for each reaction: 100 pg template DNA (0.74 fmol) with 500 pg sonified salmon sperm DNA as carrier, 25 ng (3.3 pmol) of each 5'-phosphorylated oligonucleotide, 20 ng (2.5 pmol) of each non-phosphorylated oligonucleotide, 4 U *Pfu* DNA ligase in a final volume of 20  $\mu$ l buffered ss 50-mer was used (1 fmol) as  
 20 template, in this case oligo C was also biotinylated. All reactions were performed in a thermocycler (OmniGene, MWG-Biotech, Ebersberg, Germany) with the following program: 4 minutes 92°C, 2 minutes 60°C and 25 cycles of 20 seconds 92°C, 40 seconds 60°C. Except for HPLC analysis the biotinylated ligation educt C was used. In a control  
 25 experiment the biotinylated and non-biotinylated oligonucleotides revealed the same gel electrophoretic results. The reactions were analyzed on 7.5% polyacrylamide gels. Ligation product 1 (oligo A and B) calculated mass: 15450 Da, ligation product 2 (oligo C and D) calculated mass: 15387 Da.

30 *SMART-HPLC*



-91-

Ion exchange HPLC (IE HPLC) was performed on the SMART-system (Pharmacia, Freiburg, Germany) using a Pharmacia Mono Q, PC 1.6/5 column. Eluents were buffer A (25 mM Tris-HCl, 1 mM EDTA and 0.3 M NaCl at pH 8.0) and buffer B (same as A, but 1 M NaCl). Starting with 100% A for 5 minutes at a flow rate of 50  $\mu$ l/min. a gradient was applied from 0 to 70% B in 30 minutes, then increased to 100% B in 2 minutes and held at 100% B for 5 minutes. Two pooled LCR volumes (40  $\mu$ l) performed with either wildtype or mutant template were injected.

#### 10      *Sample preparation for MALDI-TOF-MS*

Preparation of immobilized DNA: For the recording of each spectrum two LCRs (performed as described above) were pooled and diluted 1:1 with 2x B/W buffer (10 mM Tris-HCl, pH 7.5, 1 mM EDTA, 2 M NaCl). To the samples 5  $\mu$ l streptavidin DynaBeads (Dyna, Hamburg, Germany) were added, the mixture was allowed to bind with gentle shaking for 15 minutes at ambient temperature. The supernatant was removed using a Magnetic Particle Collector, MPC, (Dyna, Hamburg, Germany) and the beads were washed twice with 50  $\mu$ l of 0.7 M ammonium citrate solution (pH 8.0) (the supernatant was removed each time using the MPC). The beads were resuspended in 1  $\mu$ l of ultrapure water (MilliQ, Millipore, Bedford, Mabelow).

Combination of ultrafiltration and streptavidin DynaBeads: For the recording of spectrum two LCRs (performed as described above) were pooled, diluted 1:1 with 2x B/W buffer and concentrated with a 5000 NMWL Ultrafree-MC filter unit (Millipore, Eschborn, Germany) according to the instructions of the manufacturer. After concentration the samples were washed with 300  $\mu$ l 1x B/W buffer to streptavidin DynaB ads were added. The beads were washed once on the Ultrafree-MC filtration unit with 300  $\mu$ l of 1x B/W buffer and processed as described above. The

-92-

beads were resuspended in 30 to 50  $\mu$ l of 1x B/W buffer and transferred in a 1.5 ml Eppendorf tube. The supernatant was removed and the beads were washed twice with 50  $\mu$ l of 0.7 M ammonium citrate (pH 8.0). Finally, the beads were washed once with 30  $\mu$ l of acetone and resuspended in 1  $\mu$ l of ultrapure water. The ligation mixture after immobilization on the beads was used for MALDI-TOF-MS analysis as described below.

#### MALDI-TOF-MS

A suspension of streptavidin-coated magnetic beads with the immobilized DNA was pipetted onto the sample holder, then immediately mixed with 0.5  $\mu$ l matrix solution (0.7 M 3-hydroxypicolinic acid in 50% acetonitrile, 70 mM ammonium citrate). This mixture was dried at ambient temperature and introduced into the mass spectrometer. All spectra were taken in positive ion mode using a Finnigan MAT Vision 2000 (Finnigan MAT, Bremen, Germany), equipped with a reflectron (5 keV ion source, 20 keV postacceleration) and a nitrogen laser (337 nm). For the analysis of *Pfu* DNA ligase 0.5  $\mu$ l of the solution was mixed on the sample holder with 1  $\mu$ l of matrix solution and prepared as described above. For the analysis of unpurified LCRs 1  $\mu$ l of an LCR was mixed with 1  $\mu$ l matrix solution.

#### RESULTS

The *E. coli lacI* gene served as a simple model system to investigate the suitability of MALDI-TOF-MS as detection method for products generated in ligase chain reactions. This template system contains of an *E. coli lacI* wildtype gene in a pBluescript KII phagemid and an *E. coli lacI* gene carrying a single point mutation at bp 191 (C to T transition; SEQ ID No. 131) in the same phagemid. Four different oligonucleotides were used, which were ligated only if the *E. coli lacI* wildtype gene was present (Figure 26).

-93-

LCR conditions were optimized using *Pfu* DNA ligase to obtain at least 1 pmol ligation product in each positive reaction. The ligation reactions were analyzed by polyacrylamide gel electrophoresis (PAGE) and HPLC on the SMART system (Figures 27, 28 and 29). Figure 27 shows a PAGE of a positive LCR with wildtype template (lane 1), a negative LCR with mutant template (1 and 2) and a negative control which contains enzyme, oligonucleotides and no template but salmon sperm DNA. The gel electrophoresis clearly shows that the ligation product (50 bp) was produced only in the reaction with wildtype template; whereas neither the template carrying the point mutation nor the control reaction with salmon sperm DNA generated amplification products. In Figure 28, HPLC was used to analyze two pooled LCRs with wildtype template performed under the same conditions. The ligation product was clearly revealed. Figure 29 shows the results of a HPLC in which two pooled negative LCRs with mutant template were analyzed. These chromatograms confirm the data shown in Figure 27 and the results taken together clearly demonstrate, that the system generates ligation products in a significant amount only if the wildtype template is provided.

Appropriate control runs were performed to determine retention times of the different compounds involved in the LCR experiments. These include the four oligonucleotides (A, B, C, and D), a synthetic ds 50-mer (with the same sequence as the ligation product), the wildtype template DNA, sonicated salmon sperm DNA and the *Pfu* DNA ligase in ligation buffer.

In order to test which purification procedure should be used before a LCR reaction can be analyzed by MALDI-TOF-MS, aliquots of an unpurified LCR (Figure 30A) and aliquots of the enzyme stock solution (Figure 30B) were analyzed with MALDI-TOF-MS. It turned out that

-94-

appropriate sample preparation is absolutely necessary since all signals in the unpurified LCR correspond to signals obtained in the MALDI-TOF-MS analysis of the *Pfu* DNA ligase. The calculated mass values of oligo A and the ligation product are 7521 Da and 15450 Da, respectively. The data in Figure 30 show that the enzyme solution leads to mass signals which do interfere with the expected signals of the ligation educts and products and therefore makes an unambiguous signal assignment impossible. Furthermore, the spectra showed signals of the detergent Tween20 being part of the enzyme storage buffer which influences the crystallization behavior of the analyte/matrix mixture in an unfavorable way.

In one purification format streptavidin-coated magnetic beads were used. As was shown in a recent paper, the direct desorption of DNA immobilized by Watson-Crick base pairing to a complementary DNA fragment covalently bound to the beads is possible and the non-biotinylated strand will be desorbed exclusively (Tang *et al.* (1995) Nucleic Acids Res. 23:3126-3131). This approach in using immobilized ds DNA ensures that only the non-biotinylated strand will be desorbed. If non-immobilized ds DNA is analyzed both strands are desorbed (Tang *et al.* (1994) Rapid Comm. Mass Spectrom. 7 183-186) leading to broad signals depending on the mass difference of the two single strands. Therefore, employing this system for LCR only the non-ligated oligonucleotide A, with a calculated mass of 7521 Da, and the ligation product from oligo A and oligo B (calculated mass: 15450 Da) will be desorbed if oligo C is biotinylated at the 5'-end and immobilized on streptavidin-coated beads. This results in a simple and unambiguous identification of the LCR educts and products.

Figure 31A shows a MALDI-TOF mass spectrum obtained from two pooled LCRs (performed as described above) purified on streptavidin

-95-

DynaBeads and desorbed directly from the beads showed that the purification method used was efficient (compared with Figure 30). A signal which represents the unligated oligo A and a signal which corresponds to the ligation product could be detected. The agreement  
5 between the calculated and the experimentally found mass values is remarkable and allows an unambiguous peak assignment and accurate detection of the ligation product. In contrast, no ligation product but only oligo A could be detected in the spectrum obtained from two pooled LCRs with mutated template (Figure 31B). The specificity and selectivity  
10 of the LCR conditions and the sensitivity of the MALDI-TOF detection is further demonstrated when performing the ligation reaction in the absence of a specific template. Figure 32 shows a spectrum obtained from two pooled LCRs in which only salmon sperm DNA was used as a negative control, only oligo A could be detected, as expected.

15 While the results shown in Figure 31A can be correlated to lane 1 of the gel in Figure 27, the spectrum shown in Figure 31B is equivalent to lane 2 in Figure 27, and finally also the spectrum in Figure 32 corresponds to lane 3 in Figure 27. The results are in congruence with the HPLC analysis presented in Figures 28 and 29. While gel  
20 electrophoresis (Figure 27) and HPLC (Figures 28 and 29) reveal either an excess or almost equal amounts of ligation product over ligation educts, the analysis by MALDI-TOF mass spectrometry produces a smaller signal for the ligation product (Figure 31A).

The lower intensity of the ligation product signal could be due to  
25 different desorption/ionization efficiencies between 24- and a 50-mer. Since the  $T_m$  value of a duplex with 50 compared to 24 base pairs is significantly higher, more 24-mer could be desorbed. A reduction in signal intensity can also result from a higher degree of fragmentation in case of the longer oligonucleotides.

Regardless of the purification with streptavidin DynaBeads, Figure 32 reveals traces of Tween20 in the region around 2000 Da. Substances with a viscous consistence, negatively influence the process of crystallization and therefore can be detrimental to mass spectrometer analysis. Tween20 and also glycerol which are part of enzyme storage buffers therefore should be removed entirely prior to mass spectrometer analysis. For this reason an improved purification procedure which includes an additional ultrafiltration step prior to treatment with DynaBeads was investigated. Indeed, this sample purification resulted in a significant improvement of MALDI-TOF mass spectrometric performance.

Figure 33 shows spectra obtained from two pooled positive (Fig. 33A) and negative (Fig. 33B) LCRs, respectively. The positive reaction was performed with a chemically synthesized, single strand 50mer as template with a sequence equivalent to the ligation product of oligo C and D. Oligo C was 5'-biotinylated. Therefore the template was not detected. As expected, only the ligation product of Oligo A and B (calculated mass 15450 Da) could be desorbed from the immobilized and ligated oligo C and D. This newly generated DNA fragment is represented by the mass signal of 15448 Da in Figure 33A. Compared to Figure 32A, this spectrum clearly shows that this method of sample preparation produces signals with improved resolution and intensity.

#### EXAMPLE 7

Mutation detection by solid phase oligo base extension of a primer and analysis by MALDI-TOF mass spectrometry (Primer Oligo Base Extension = Probe)

##### *Summary*

The solid-phase oligo base extension method detects point mutations and small deletions as well as small insertions in amplified DNA. The method is based on the extension of a detection primer that

-97-

anneals adjacent to a variable nucleotide position on an affinity-captured amplified template, using a DNA polymerase, a mixture of three dNTPs, and the missing one dideoxy nucleotide. The resulting products are evaluated and resolved by MALDI-TOF mass spectrometry without  
5 further labeling procedures. The aim of the following experiment was to determine mutant and wildtype alleles in a fast and reliable manner.

*Description of the experiment*

The method used a single detection primer followed by a oligonucleotide extension step to give products differing in length by  
10 some bases specific for mutant or wildtype alleles which can be easily resolved by MALDI-TOF mass spectrometry. The method is described by using as example the exon 10 of the CFTR-gene. Exon 10 of this gene bears the most common mutation in many ethnic groups ( $\Delta F508$ ) that leads in the homozygous state to the clinical phenotype of cystic fibrosis.

**15 MATERIALS AND METHODS**

*Genomic DNA*

Genomic DNA were obtained from healthy individuals, individuals homozygous or heterozygous for the  $\Delta F508$  mutation, and one individual heterozygous for the 1506S mutation. The wildtype and mutant alleles  
20 were confirmed by standard Sanger sequencing.

*PCR amplification of exon 10 of the CFTR gene*

The primers for PCR amplification were CFEx10-F (5'-GCAAGTGAATCCTGAGCGTG-3' (SEQ ID No. 13) located in intron 9 and biotinylated) and CFEx10-R (5'-GTGTGAAGGGCGTG-3' SEQ ID No. 14)  
25 located in intron 10). Primers were used in a concentration of 8 pmol. Taq-polymerase including 10x buffer were purchased from Boehringer-Mannheim and dNTPs were obtained from Pharmacia. The total reaction volume was 50  $\mu$ l. Cycling conditions for PCR were initially

-98-

5 min. at 95°C, followed by 1 min. at 94°C, 45 sec at 53°C, and 30 sec at 72°C for 40 cycles with a final extension time of 5 min at 72°C.

*Purification of the amplified products*

Amplification products were purified by using Qiagen's PCR  
5 purification kit (No. 28106) according to manufacturer's instructions.

The elution of the purified products from the column was done in 50 µl TE-buffer (10mM Tris, 1 mM EDTA, pH 7,5).

*Affinity-capture and denaturation of the double stranded DNA*

10 10 µL aliquots of the purified amplified product were transferred to one well of a streptavidin-coated microtiter plate (No. 1645684 Boehringer-Mannheim or No. 95029262 Labsystems). Subsequently, 10 µl incubation buffer (80 mM sodium phosphate, 400 mM NaCl, 0,4% Tween20, pH 7,5) and 30 µl water were added. After incubation for 1 hour at room temperature the wells were washed three times with 200µl  
15 washing buffer (40 mM Tris, 1 mM EDTA, 50 mM NaCl, 0.1% Tween 20, pH 8.8). To denature the double stranded DNA the wells were treated with 100 µl of a 50 mM NaOH solution for 3 min and the wells washed three times with 200 µl washing buffer.

*Oligo base extension reaction*

20 The annealing of 25 pmol detection primer (CF508: 5'-CTATATTCATCATAGGAAACACCA-3' (SEQ ID No. 15) was performed in 50 µl annealing buffer (20 mM Tris, 10 mM KCl, 10 mM (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, 2 mM MgSO<sub>2</sub>, 1% Triton X-100, pH 8) at 50°C for 10 min. The wells were washed three times with 200 µl washing buffer and once in 200 µl  
25 TE buffer. The extension reaction was performed by using some components of the DNA sequencing kit from USB (No. 70770) and dNTPs or ddNTPs from Pharmacia. The total reaction volume was 45 µl, containing of 21 µl water, 6 µl Sequenase-buffer, 3 µl 10 mM DTT solution, 4,5 µl, 0,5 mM of three dNTPs, 4,5 µl, 2 mM the missing one



-99-

ddNTP, 5,5  $\mu$ l glycerol enzyme dilution buffer, 0,25  $\mu$ l Sequenase 2.0, and 0,25 pyrophosphatase. The reaction was pipetted on ice and then incubated for 15 min at room temperature and for 5 min at 37°C.

Hence, the wells were washed three times with 200  $\mu$ l washing buffer and once with 60  $\mu$ l of a 70 mM  $\text{NH}_4$ -Citrate solution.

*Denaturation and precipitation of the extended primer*

The extended primer was denatured in 50  $\mu$ l 10%-DMSO (dimethylsulfoxide) in water at 80°C for 10 min. For precipitation, 10  $\mu$ l  $\text{NH}_4$ -Acetate (pH 6.5), 0,5  $\mu$ l glycogen (10 mg/ml water, Sigma No.

10 G1765), and 100  $\mu$ l absolute ethanol were added to the supernatant and incubated for 1 hour at room temperature. After centrifugation at 13.000 g for 10 min the pellet was washed in 70% ethanol and resuspended in 1  $\mu$ l 18 Mohm/cm  $\text{H}_2\text{O}$  water.

*Sample preparation and analysis on MALDI-TOF mass*

15 *spectrometry*

Sample preparation was performed by mixing 0,3  $\mu$ l of each of matrix solution (0.7 M 3-hydroxypicolinic acid, 0.07 M dibasic ammonium citrate in 1:1  $\text{H}_2\text{O}:\text{CH}_3\text{CN}$ ) and of resuspended DNA/glycogen pellet on a sample target and allowed to air dry. Up to 20 samples were spotted on a probe target disk for introduction into the source region of an unmodified Thermo Bioanalysis (formerly Finnigan) Visions 2000 MALDI-TOF operated in reflectron mode with 5 and 20 kV on the target and conversion dynode, respectively. Theoretical average molecular mass ( $M_r(\text{calc})$ ) were calculated from atomic compositions; reported experimental  $M_r$  ( $M_r(\text{exp})$ ) values are those of the singly-protonated form, determined using external calibration.

**RESULTS**

The aim of the experiment was to develop a fast and reliable method independent of exact stringencies for mutation detection that

-100-

leads to high quality and high throughput in the diagnosis of genetic diseases. Therefore a special kind of DNA sequencing (oligo base extension of one mutation detection primer) was combined with the evaluation of the resulting mini-sequencing products by matrix-assisted  
5 laser desorption ionization (MALDI) mass spectrometry (MS). The time-of-flight (TOF) reflectron arrangement was chosen as a possible mass measurement system. To prove this hypothesis, the examination was performed with exon 10 of the CFTR-gene, in which some mutations could lead to the clinical phenotype of cystic fibrosis, the most common  
10 monogenetic disease in the Caucasian population.

The schematic presentation as given in Figure 34 shows the expected short sequencing products with the theoretically calculated molecular mass of the wildtype and various mutations of exon 10 of the CFTR-gene (SEQ ID No. 132). The short sequencing products were  
15 produced using either ddTTP (Figure 34A; SEQ ID Nos. 133-135) or ddCTP (Figure 34B; SEQ ID Nos. 136-139) to introduce a definitive sequence related stop in the nascent DNA strand. The MALDI-TOF-MS spectra of healthy, mutation heterozygous, and mutation homozygous individuals are presented in Figure 35. All samples were confirmed by  
20 standard Sanger sequencing which showed no discrepancy in comparison to the mass spec analysis. The accuracy of the experimental measurements of the various molecular masses was within a range of minus 21.8 and plus 87.1 dalton (Da) to the range expected. This allows a definitive interpretation of the results in each case. A further  
25 advantage of this procedure is the unambiguous detection of the  $\Delta$ I507 mutation. In the ddTTP reaction, the wildtype allele would be detected, whereas in the ddCTP reaction the three base pair deletion would be disclosed.

-101-

- The method described is highly suitable for the detection of single point mutations or microlesions of DNA. Careful choice of the mutation detection primers will open the window of multiplexing and lead to a high throughput including high quality in genetic diagnosis without any need
- 5 for exact stringencies necessary in comparable allele-specific procedures. Because of the uniqueness of the genetic information, the oligo base extension of mutation detection primer is applicable in each disease gene or polymorphic region in the genome like variable number of tandem repeats (VNTR) or other single nucleotide polymorphisms (e.g.,
- 10 apolipoprotein E gene), as also described here.

#### EXAMPLE 8

Detection of Polymerase Chain Reaction Products Containing 7-Deazapurine Moieties with Matrix-Assisted Laser Desorption/Ionization Time-of-Flight (MALDI-TOF) Mass Spectrometry

#### 15 MATERIALS AND METHODS

##### *Nucleic acid amplifications*

- The following oligodeoxynucleotide primers were either synthesized according to standard phosphoramidite chemistry (Sinha, N.D., et al., (1983) *Tetrahedron Let.* Vol. 24, Pp. 5843-5846; Sinha,
- 20 N.D., et al., (1984) *Nucleic Acids Res.*, Vol. 12, Pp. 4539-4557) on a MilliGen 7500 DNA synthesizer (Millipore, Bedford, MA, USA) in 200 nmol scales or purchased from MWG-Biotech (Ebersberg, Germany, primer 3) and Biometra (Goettingen, Germany, primers 6-7).
- primer 1: 5'-GTCACCCTCGACCTGCAG (SEQ ID NO. 16);
- 25 primer 2: 5'-TTGTAAAACGACGGCCAGT (SEQ ID NO. 17);
- primer 3: 5'-CTTCCACCGCGATGTTGA (SEQ ID NO. 18);
- primer 4: 5'-CAGGAAACAGCTATGAC (SEQ ID NO. 19);
- primer 5: 5'-GTAAAACGACGGCCAGT (SEQ ID NO. 20);
- primer 6: 5'-GTCACCCTCGACCTGCAGC (g: RiboG) (SEQ ID NO. 21);

-102-

primer 7: 5'-GTTGTAAAACGAGGGCCAgT (g: RiboG) (SEQ ID NO. 22);

The 99-mer (SEQ ID No. 141) and 200-mer DNA strands (SEQ ID No. 140; modified and unmodified) as well as the ribo- and

- 5 7-deaza-modified 100-mer were amplified from pRFc1 DNA (10 ng, generously supplied by S. Feyerabend, University of Hamburg) in 100  $\mu$ L reaction volume containing 10 mmol/L KCl, 10 mmol/L  $(\text{NH}_4)_2\text{SO}_4$ , 20 mmol/L Tris HCl (pH 8.8), 2 mmol/L  $\text{MgSO}_4$ , (exo(-) *Pseudococcus furiosus* (*Pfu*) -Buffer, Pharmacia, Freiburg, Germany), 0.2 mmol/L each
- 10 dNTP (Pharmacia, Freiburg, Germany), 1  $\mu$ mol/L of each primer and 1 unit of exo(-)*Pfu* DNA polymerase (Stratagene, Heidelberg, Germany). For the 99-mer primers 1 and 2, for the 200-mer primers 1 and 3 and for the 100-mer primers 6 and 7 were used. To obtain 7-deazapurine modified nucleic acids, during PCR-amplification dATP and dGTP were
- 15 replaced with 7-deaza-dATP and 7-deaza- dGTP. The reaction was performed in a thermal cycler (OmniGene, MWG-Biotech, Ebersberg, Germany) using the cycle: denaturation at 95°C for 1 min., annealing at 51°C for 1 min. and extension at 72°C for 1 min. For all PCRs the number of reaction cycles was 30. The reaction was allowed to extend
- 20 for additional 10 min. at 72°C after the last cycle.

- The 103-mer DNA strands (modified and unmodified; SEQ ID No. 245) were amplified from M13mp18 RFI DNA (100 ng, Pharmacia, Freiburg, Germany) in 100  $\mu$ L reaction volume. using primers 4 and 5 all other concentrations were unchanged. The reaction was performed
- 25 using the cycle: denaturation at 95°C for 1 min., annealing at 40°C for 1 min. and extension at 72°C for 1 min. After 30 cycles for the unmodified and 40 cycles for the modified 103-mer respectively, the samples were incubated for additional 10 min. at 72°C.

*Synthesis of 5'-[<sup>32</sup>P]-labeled PCR-primers*

-103-

Primers 1 and 4 were 5'-[<sup>32</sup>P]-labeled employing T4-polynucleotidkinase (Epicentre Technologies) and ( $\gamma$ -<sup>32</sup>P)-ATP. (BLU/NGG/502A, Dupont, Germany) according to the protocols of the manufacturer. The reactions were performed substituting 10% of primer 1 and 4 in PCR with the labeled primers under otherwise unchanged reaction-conditions. The amplified DNAs were separated by gel electrophoresis on a 10% polyacrylamide gel. The appropriate bands were excised and counted on a Packard TRI-CARB 460C liquid scintillation system (Packard, CT, USA).

10 *Primer-cleavage from ribo-modified PCR-product*

The amplified DNA was purified using Ultrafree-MC filter units (30,000 NMWL), it was then redissolved in 100  $\mu$ l of 0.2 mol/L NaOH and heated at 95°C for 25 minutes. The solution was then acidified with HCl (1 mol/L) and further purified for MALDI-TOF analysis employing Ultrafree-MC filter units (10,000 NMWL) as described below.

*Purification of amplified products*

All samples were purified and concentrated using Ultrafree-MC units 30000 NMWL (Millipore, Eschborn, Germany) according to the manufacturer's description. After lyophilization, amplified products were redissolved in 5  $\mu$ L (3  $\mu$ L for the 200-mer) of ultrapure water. This analyte solution was directly used for MALDI-TOF measurements.

*MALDI-TOF MS*

Aliquots of 0.5  $\mu$ L of analyte solution and 0.5  $\mu$ L of matrix solution (0.7 mol/L 3-HPA and 0.07 mol/L ammonium citrate in acetonitrile/water (1:1, v/v)) were mixed on a flat metallic sample support. After drying at ambient temperature the sample was introduced into the mass spectrometer for analysis. The MALDI-TOF mass spectrometer used was a Finnigan MAT Vision 2000 (Finnigan MAT, Bremen, Germany). Spectra were recorded in the positive ion reflector mode with a 5 keV ion source

-104-

and 20 keV postacceleration. The instrument was equipped with a nitrogen laser (337 nm wavelength). The vacuum of the system was  $3\text{--}4 \times 10^{-8}$  hPa in the analyzer region and  $1\text{--}4 \times 10^{-7}$  hPa in the source region. Spectra of modified and unmodified DNA samples were obtained with the same relative laser power; external calibration was performed with a mixture of synthetic oligodeoxynucleotides (7-to 50-mer).

## RESULTS AND DISCUSSION

### *Enzymatic synthesis of 7-deazapurine nucleotide containing nucleic acids by PCR*

10 In order to demonstrate the feasibility of MALDI-TOF MS for the rapid, gel-free analysis of short amplified products and to investigate the effect of 7-deazapurine modification of nucleic acids under MALDI-TOF conditions, two different primer-template systems were used to synthesize DNA fragments. Sequences are displayed in Figures 36 and 15 37. While the two single strands of the 103-mer amplified product had nearly equal masses ( $\Delta m = 8$  u), the two single strands of the 99-mer differed by 526 u. Considering that 7-deaza purine nucleotide building blocks for chemical DNA synthesis are approximately 160 times more expensive than regular ones (Product Information, Glen Research Corporation, Sterling, VA) and their application in standard  $\beta$ - 20 cyano-phosphoramidite chemistry is not trivial (Product Information, Glen Research Corporation, Sterling, VA; Schneider *et al.* (1995) *Nucl. Acids Res.* 23:1570) the cost of 7-deaza purine modified primers would be very high. Therefore, to increase the applicability and scope of the method, 25 all PCRs were performed using unmodified oligonucleotide primers which are routinely available. Substituting dATP and dGTP by  $c^7$ -dATP and  $c^7$ -dGTP in polymerase chain reaction led to products containing approximately 80% 7-deaza-purine modified nucleosides for the 99-mer

-105-

and 103-mer; and about 90% for the 200-mer, respectively. Table II shows the base composition of all PCR products.

TABLE II:

**5 Base composition of the 99-mer, 103-mer and 200-mer PCR amplification products (unmodified and 7-deaza purine modified)**

	DNA-fragments <sup>1</sup>	C	T	A	G	c <sup>7</sup> -deaza-A	c <sup>7</sup> -deaza-G	rel. mod.2
	200-mers	54	34	56	56	-	-	-
	modified 200-mer s	54	34	6	5	50	51	90%
10	200-mer a	56	56	34	54	-	-	-
	modified 200-mer a	56	56	3	4	31	50	92%
	103-mer s	28	23	24	28	-	-	-
	modified 103-mer s	28	23	6	5	18	23	79%
	103-mer a	28	24	23	28	-	-	-
15	modified 103-mer a	28	24	7	4	16	24	78%
	99-mer s	34	21	24	20	-	-	-
	modified 99-mer s	34	21	6	5	18	15	75%
	99-mer a	20	24	21	34	-	-	-
20	modified 99-mer a	20	24	3	4	18	30	87%

<sup>1</sup> "s" and "a" describe "sense" and "antisense" strands of the double-stranded amplified product.

<sup>2</sup> indicates relative modification as percentage of 7-deaza purine modified nucleotides of total amount of purine nucleotides.

25

It remained to be determined whether 80-90% 7-deaza-purine modification is sufficient for accurate mass spectrometer detection. It was therefore important to determine whether all purine nucleotides could be substituted during the enzymatic amplification step. This was not trivial since it had been shown that c<sup>7</sup>-dATP cannot fully replace dATP in PCR if *Taq* DNA polymerase is employed (Seela, F. and A. Rölling (1992) *Nucleic Acids Res.*, 20,55-61). Fortunately it was found

30

-106-

that *exo(-)Pfu* DNA polymerase indeed could accept  $c^7$ -dATP and  $c^7$ -dGTP in the absence of unmodified purine nucleoside triphosphates. The incorporation was less efficient leading to a lower yield of amplified product (Figure 38).

- 5 To verify these results, the amplifications with [ $^{32}$ P]-labeled primers were repeated. The autoradiogram (Figure 39) clearly shows lower yields for the modified PCR-products. The bands were excised from the gel and counted. For all amplified products the yield of the modified nucleic acids was about 50%, referring to the corresponding unmodified
- 10 amplification product. Further experiments showed that *exo(-)DeepVent* and *Vent* DNA polymerase were able to incorporate  $c^7$ -dATP and  $c^7$ -dGTP during PCR as well. The overall performance, however, turned out to be best for the *exo(-)Pfu* DNA polymerase giving least side products during amplification. Using all three polymerases, it was found that such PCRs
- 15 employing  $c^7$ -dATP and  $c^7$ -dGTP instead of their isosteres showed less side-reactions giving a cleaner PCR-product. Decreased occurrence of amplification side products may be explained by a reduction of primer mismatches due to a long template which is synthesized during PCR. Decreased melting point for DNA duplexes containing 7-deaza-purine
- 20 have been described (Mizusawa, S. et al., (1986) *Nucleic Acids Res.*, 14, 1319-1324). In addition to the three polymerases specified above (*exo(-)Deep Vent* DNA polymerase, *5Vent* DNA polymerase and *exo(-) (Pfu)* DNA polymerase), it is anticipated that other polymerases, such as the Large Klenow fragment of *E.coli* DNA polymerase, *Sequenase*, *Taq* DNA
- 25 polymerase and *U AmpliTaq* DNA polymerase can be used. In addition, where RNA is the template, RNA polymerases, such as the SP6 or the T7 RNA polymerase, must be used.



-107-

***MALDI-TOF mass spectrometry of modified and unmodified amplified products.***

The 99-mer, 103-mer and 200-mer amplified products were analyzed by MALDI-TOF MS. Based on past experience, it was known that the degree of depurination depends on the laser energy used for desorption and ionization of the analyte. Since the influence of 7-deazapurine modification on fragmentation due to depurination was to be investigated, all spectra were measured at the same relative laser energy.

Figures 40a and 40b show the mass spectra of the modified and unmodified 103-mer nucleic acids. In case of the modified 103-mer, fragmentation causes a broad  $(M+H)^+$  signal. The maximum of the peak is shifted to lower masses so that the assigned mass represents a mean value of  $(M+H)^+$  signal and signals of fragmented ions, rather than the  $(M+H)^+$  signal itself. Although the modified 103-mer still contains about 20% A and G from the oligonucleotide primers, it shows less fragmentation which is featured by much more narrow and symmetric signals. Especially peak tailing on the lower mass side due to depurination, is substantially reduced. Hence, the difference between measured and calculated mass is strongly reduced although it is still below the expected mass. For the unmodified sample a  $(M+H)^+$  signal of 31670 was observed, which is a 97 u or 0.3% difference to the calculated mass. While, in case of the modified sample this mass difference diminished to 10 u or 0.03% (31713 u found, 31723 u calculated). These observations are verified by a significant increase in mass resolution of the  $(M+H)^+$  signal of the two signal strands ( $n/\Delta m = 67$  as opposed to 18 for the unmodified sample with  $\Delta m =$  full width at half maximum, fwhm). Because of the low mass difference between the two single strands (8 u) their individual signals were not resolved.

-108-

With the results of the 99 base pair DNA fragments the effects of increased mass resolution for 7-deazapurine containing DNA becomes even more evident. The two single strands in the unmodified sample were not resolved even though the mass difference between the two strands of the amplified product was very high with 526 u due to unequal distribution of purines and pyrimidines (figure 41a). In contrast to this, the modified DNA showed distinct peaks for the two single strands (figure 41b) which demonstrates the superiority of this approach for the determination of molecular weights to gel electrophoretic methods even more profound. Although base line resolution was not obtained the individual masses were able to be assigned with an accuracy of 0.1%:  $\Delta m = 27$  u for the lighter (calc. mass = 30224 u) and  $\Delta m = 14$  u for the heavier strand (calc. mass = 30750 u). Again, it was found that the full width at half maximum was substantially decreased for the 7-deazapurine containing sample.

In case the 99-mer and 103-mer, the 7-deazapurine containing nucleic acids seem to give higher sensitivity despite the fact that they still contain about 20% unmodified purine nucleotides. To get comparable signal-to-noise ratio at similar intensities for the  $(M+H)^+$  signals, the unmodified 99-mer required 20 laser shots in contrast to 12 for the modified one and the 103-mer required 12 shots for the unmodified sample as opposed to three for the 7-deazapurine nucleoside-containing amplified product.

Comparing the spectra of the modified and unmodified 200-mer amplicons, improved mass resolution was again found for the 7-deazapurine containing sample as well as increased signal intensities (Figures 42A and 42B). While the signal of the single strands predominates in the spectrum of the modified sample the DNA-duplex

-109-

and dimers of the single strands gave the strongest signal for the unmodified sample.

A complete 7-deaza purine modification of nucleic acids may be achieved either using modified primers in PCR or cleaving the unmodified  
5- primers from the partially modified amplified product. Since disadvantages are associated with modified primers, as described above, a 100-mer was synthesized using primers with a ribo-modification. The primers were cleaved hydrolytically with NaOH according to a method developed earlier in our laboratory (Koester, H. *et al.*, *Z Physiol. Chem.*,  
10 359, 1570-1589). Figures 43A and 43B display the spectra of the amplified product before and after primer cleavage. Figure 43b shows that the hydrolysis was successful: The hydrolyzed amplified product as well as the two released primers could be detected together with a small signal from residual uncleaved 100-mer. This procedure is especially  
15 useful for the MALDI-TOF analysis of very short PCR-products since the share of unmodified purines originating from the primer increases with decreasing length of the amplified sequence.

The remarkable properties of 7-deazapurine modified nucleic acids can be explained by either more effective desorption and/or ionization,  
20 increased ion stability and/or a lower denaturation energy of the double stranded purine modified nucleic acid. The exchange of the N-7 for a methyl group results in the loss of one acceptor for a hydrogen bond which influences the ability of the nucleic acid to form secondary structures due to non-Watson-Crick base pairing (Seela, F. and A. Kehne  
25 (1987) *Biochemistry*, 26, 2232-2238.). In addition to this the aromatic system of 7-deazapurine has a lower electron density that weakens Watson-Crick base pairing resulting in a decreased melting point (Mizusawa, S. *et al.*, (1986) *Nucleic Acids Res.*, 14, 1319-1324) of the double-strand. This effect may decrease the energy needed for

-110-

denaturation of the duplex in the MALDI process. These aspects as well as the loss of a site which probably will carry a positive charge on the N-7 nitrogen renders the 7-deazapurine modified nucleic acid less polar and may promote the effectiveness of desorption.

- 5        Because of the absence of N-7 as proton acceptor and the decreased polarization of the C-N bond in 7-deazapurine nucleosides depurination following the mechanisms established for hydrolysis in solution is prevented. Although a direct correlation of reactions in solution and in the gas phase is problematic, less fragmentation due to
- 10    depurination of the modified nucleic acids can be expected in the MALDI process. Depurination may either be accompanied by loss of charge which decreases the total yield of charged species or it may produce charged fragmentation products which decreases the intensity of the non fragmented molecular ion signal.
- 15        The observation of increased sensitivity and decreased peak tailing of the  $(M+H)^+$  signals on the lower mass side due to decreased fragmentation of the 7-deazapurine containing samples indicate that the N-7 atom indeed is essential for the mechanism of depurination in the MALDI-TOF process. In conclusion, 7-deazapurine containing nucleic
- 20    acids show distinctly increased ion-stability and sensitivity under MALDI-TOF conditions and therefore provide for higher mass accuracy and mass resolution.

#### EXAMPLE 9

##### Solid Phase Sequencing and Mass Spectrometer Detection

#### 25 MATERIALS AND METHODS

Oligonucleotides were purchased from Operon Technologies (Alameda, CA) in an unpurified form. Sequencing reactions were performed on a solid surface using reagents from the sequencing kit for Sequenase Version 2.0 (Amersham, Arlington Heights, Illinois).

-111-

*Sequencing a 39-mer target*

## Sequencing complex:

	SEQUENCE	SEQ ID NO.
5	5'-TCTGGCCTGGTGCAGGGCCTATTGTAGTTGTGACGTACA-(A <sup>9</sup> )-3'	23
	5'-TGTACGTCACAACT-3' (PNA 16/DNA)	24

In order to perform solid-phase DNA sequencing, template strand DNA11683 was 3'-biotinylated by terminal deoxynucleotidyl transferase.

- 10 A 30  $\mu$ l reaction, containing 60 pmol of DNA11683, 1.3 nmol of biotin 14-dATP (GIBCO BRL, Grand Island, NY), 30 units of terminal transferase (Amersham, Arlington Heights, Illinois), and 1x reaction buffer (supplied with enzyme), was incubated at 37°C for 1 hour. The reaction was stopped by heat inactivation of the terminal transferase at 70°C for 10
- 15 min. The resulting product was desalted by passing through a TE-10 spin column (Clontech). More than one molecules of biotin-14-dATP could be added to the 3'-end of DNA11683. The biotinylated DNA11683 was incubated with 0.3 mg of Dynal streptavidin beads in 30  $\mu$ l 1x binding and washing buffer at ambient temperature for 30 min.
- 20 The beads were washed twice with TE and redissolved in 30  $\mu$ l TE, 10  $\mu$ l aliquot (containing 0.1 mg of beads) was used for sequencing reactions.

The 0.1 mg beads from previous step were resuspended in a 10  $\mu$ l volume containing 2  $\mu$ l of 5x Sequenase buffer (200 mM Tris-HCl, pH 7.5, 100 mM MgCl<sub>2</sub>, and 250 mM NaCl) from the Sequenase kit and 5

25 pmol of corresponding primer PNA 16/DNA. The annealing mixture was heated to 70°C and allowed to cool slowly to room temperature over a 20-30 min time period. Then 1  $\mu$ l 0.1 M dithiothreitol solution, 1  $\mu$ l Mn buffer (0.15 M sodium isocitrate and 0.1 M MgCl<sub>2</sub>), and 2  $\mu$ l of diluted Sequenase (3.25 units) were added. The reaction mixture was divided

**-112-**

into four aliquots of 3  $\mu$ l each and mixed with termination mixes (each contains of 3  $\mu$ l of the appropriate termination mix: 32  $\mu$ M c7dATP, 32  $\mu$ M dCTP, 32  $\mu$ M c7dGTP, 32  $\mu$ M dTTP and 3.2  $\mu$ M of one of the four ddTNPs, in 50 mM NaCl). The reaction mixtures were incubated at 37°C for 2 min. After the completion of extension, the beads were precipitated and the supernatant was removed. The beads were washed twice and resuspended in TE and kept at 4°C.

### Sequencing a 78-mer target

**Sequencing complex:**

**10** 5'-AAGATCTGACCAGGGATTTCGGTTAGCGTGACTGCTGCTGCTGCT  
GCTGCTGGATGATCCGACGCATCAGATCTGG-(A<sup>b</sup>)<sub>n</sub>-3'(SEQ ID NO. 25)  
(TNR.PLASM2)

5'-CTGATGCGTCGGATCATC-3' (CM1) (SEQ ID NO. 26)

The target TNR.PLASM2 was biotinylated and sequenced using  
15 procedures similar to those described in previous section (sequencing a  
39-mer target).

### Sequencing a 15-mer target with partially duplex probe

### Sequencing complex:

**6'-F-GATGATCCGACGCATCACAGCTC<sup>3'</sup> (SEQ ID No. 27)**

20 5'-TCGGTCCAAGAGCTGTGATGCGTCGGATCATC-b-3' (SEQ ID No. 28)

CM1B3B was immobilized on Dynabeads M280 with streptavidin (Dyna, Norway) by incubating 60 pmol of CM1B3B with 0.3 magnetic beads in 30  $\mu$ l 1 M NaCl and TE (1x binding and washing buffer) at room temperature for 30 min. The beads were washed twice with TE and redissolved in 30  $\mu$ l TE, 10 or 20  $\mu$ l aliquot (containing 0.1 or 0.2 mg of beads respectively) was used for sequencing reactions.

-113-

The duplex was formed by annealing corresponding aliquot of beads from previous step with 10 pmol of DF11a5F (or 20 pmol of DF11a5F for 0.2 mg of beads) in a 9  $\mu$ l volume containing 2  $\mu$ l of 5x Sequenase buffer (200 mM Tris-HCl, pH 7.5, 100 mM MgCl<sub>2</sub>, and 250 mM NaCl) from the Sequenase kit. The annealing mixture was heated to 65°C and allowed to cool slowly to 37°C over a 20-30 min time period. The duplex primer was then mixed with 10 pmol of TS10 (20 pmol of TS10 for 0.2 mg of beads) in 1  $\mu$ l volume, and the resulting mixture was further incubated at 37°C for 5 min, room temperature for 5-10 min.

10 Then 1  $\mu$ l 0.1 M dithiothreitol solution, 1  $\mu$ l Mn buffer (0.15 M sodium isocitrate and 0.1 M MnCl<sub>2</sub>), and 2  $\mu$ l of diluted Sequenase (3.25 units) were added. The reaction mixture was divided into four aliquots of 3  $\mu$ l each and mixed with termination mixes (each contains of 4  $\mu$ l of the appropriate termination mix: 16  $\mu$ M dATP, 16  $\mu$ M dCTP, 16  $\mu$ M dGTP, 16  $\mu$ M dTTP and 1.6  $\mu$ M of one of the four ddNTPs, in 50 mM NaCl). The

15 reaction mixtures were incubated at room temperature for 5 min, and 37°C for 5 min. After the completion of extension, the beads were precipitated and the supernatant was removed. The beads were resuspended in 20  $\mu$ l TE and kept at 4°C. An aliquot of 2  $\mu$ l (out of 20  $\mu$ l) from each tube was taken and mixed with 8  $\mu$ l of formamide, the

20 resulting samples were denatured at 90-95°C for 5 min and 2  $\mu$ l (out of 10  $\mu$ l total) was applied to an ALF DNA sequencer (Pharmacia, Piscataway, NJ) using a 10% polyacrylamide gel containing 7 M urea and 0.6x TBE. The remaining aliquot was used for MALDI-TOF MS

25 analysis.

-114-

*MALDI sample preparation and instrumentation*

Before MALDI analysis, the sequencing ladder loaded magnetic beads were washed twice using 50 mM ammonium citrate and resuspended in 0.5  $\mu$ l pure water. The suspension was then loaded onto the sample target of the mass spectrometer and 0.5  $\mu$ l of saturated matrix solution (3-hydroxypicolinic acid (HPA): ammonium citrate = 10:1 mole ratio in 50% acetonitrile) was added. The mixture was allowed to dry prior to mass spectrometer analysis.

The reflectron TOFMS mass spectrometer (Vision 2000, Finnigan MAT, Bremen, Germany) was used for analysis. 5 kV was applied in the ion source and 20 kV was applied for postacceleration. All spectra were taken in the positive ion mode and a nitrogen laser was used. Normally, each spectrum was averaged for more than 100 shots and a standard 25-point smoothing was applied.

**15 RESULTS AND DISCUSSION***Conventional solid-phase sequencing*

In conventional sequencing methods, a primer is directly annealed to the template and then extended and terminated in a Sanger dideoxy sequencing. Normally, a biotinylated primer is used and the sequencing ladders are captured by streptavidin-coated magnetic beads. After washing, the products are eluted from the beads using EDTA and formamide. Previous findings indicated that only the annealed strand of a duplex is desorbed and the immobilized strand remains on the beads. Therefore, it is advantageous to immobilize the template and anneal the primer. After the sequencing reaction and washing, the beads with the immobilized template and annealed sequencing ladder can be loaded directly onto the mass spectrometer target and mix with matrix. In MALDI, only the annealed sequencing ladder will be desorbed and ionized, and the immobilized template will remain on the target.



-115-

A 39-mer template (SEQ ID No. 23) was first biotinylated at the 3'-end by adding biotin-14-dATP with terminal transferase. More than one biotin-14-dATP molecule could be added by the enzyme. Since the template was immobilized and remained on the beads during MALDI, the number of biotin-14-dATP would not affect the mass spectra. A 14-mer primer (SEQ ID No. 24) was used for the solid-state sequencing to generate DNA fragments 3-27 below (SEQ ID Nos. 142-166). MALDI-TOF mass spectra of the four sequencing ladders are shown in Figure 44 and the expected theoretical values are shown in Table III.

-116-

TABLE III

1	5'-TCTGGCCTGGTGCAGGGCCTATTGTAGTTGTGACGTACA-(A <sup>8</sup> ) <sub>n</sub> -3'
2	3'-TCAACACTGCATGT-5'
3	3'-ATCAACACTGCATGT-5'
5	4 3'-CATCAACACTGCATGT-5'
5	5 3'-ACATCAACACTGCATGT-5'
6	3'-AACATCAACACTGCATGT-5'
7	3'-TAACATCAACACTGCATGT-5'
8	3'-ATAACATCAACACTGCATGT-5'
10	9 3'-GATAACATCAACACTGCATGT-5'
10	10 3'-GGATAACATCAACACTGCATGT-5'
11	3'-CGGATAACATCAACACTGCATGT-5'
12	3'-CCGGATAACATCAACACTGCATGT-5'
13	3'-CCCGGATAACATCAACACTGCATGT-5'
15	14 3'-TCCCGGATAACATCAACACTGCATGT-5'
15	15 3'-GTCCCGGATAACATCAACACTGCATGT-5'
16	3'-CGTCCCGGATAACATCAACACTGCATGT-5'
17	3'-ACGTCCCGGATAACATCAACACTGCATGT-5'
18	3'-CACGTCCCGGATAACATCAACACTGCATGT-5'
20	19 3'-CCACGTCCCGGATAACATCAACACTGCATGT-5'
20	20 3'-ACCACGTCCCGGATAACATCAACACTGCATGT-5'
21	3'-GACCACGTCCCGGATAACATCAACACTGCATGT-5'
22	3'-GGACCACGTCCCGGATAACATCAACACTGCATGT-5'
23	3'-CGGACCACGTCCCGGATAACATCAACACTGCATGT-5'
25	24 3'-CCGGACCACGTCCCGGATAACATCAACACTGCATGT-5'
25	25 3'-ACCGGACCACGTCCCGGATAACATCAACACTGCATGT-5'
26	3'-GACCGGACCACGTCCCGGATAACATCAACACTGCATGT-5'
27	3'-AGACCGGACCACGTCCCGGATAACATCAACACTGCATGT-5'

-117-

TABLE III (Continued)

	A-reaction	C-reaction	G-reaction	T-reaction
1.				
2.	4223.8	4223.8	4223.8	4223.8
5 3.	4521.1			
4.		4809.2		
5.	5133.4			
6.	5434.6			
7.				5737.8
10 8.	6051.1			
9.			6379.2	
10.			6704.4	
11.		6995.6		
12.		7284.8		
15 13.		7574.0		
14.				7878.2
15.			8207.4	
16.		8495.6		
17.	8808.8			
20 18.		9097.0		
19.		9386.2		
20.	9699.4			
21.			10027.6	
22.			10355.8	
25 23.		10644.0		
24.		10933.2		
25.	11246.4			

-118-

	A-reaction	C-reaction	G-reaction	T-reaction
26.			11574.8	
27.	11886.8			

The sequencing reaction produced a relatively homogenous ladder, and the full-length sequence was determined easily. One peak around 5150 appeared in all reactions are not identified. A possible explanation is that a small portion of the template formed some kind of secondary structure, such as a loop, which hindered sequenase extension. Mis-incorporation is of minor importance, since the intensity of these peaks were much lower than that of the sequencing ladders. Although 7-deaza purines were used in the sequencing reaction, which could stabilize the N-glycosidic bond and prevent depurination, minor base losses were still observed since the primer was not substituted by 7-deazapurines. The full length ladder, with a ddA at the 3' end, appeared in the A reaction with an apparent mass of 11899.8. A more intense peak of 12333 appeared in all four reactions and is likely due to an addition of an extra nucleotide by the Sequenase enzyme.

The same technique could be used to sequence longer DNA fragments. A 78-mer template containing a CTG repeat (SEQ ID No. 25) was 3'-biotinylated by adding biotin-14-dATP with terminal transferase. An 18-mer primer (SEQ ID No. 26) was annealed right outside the CTG repeat so that the repeat could be sequenced immediately after primer extension. The four reactions were washed and analyzed by MALDI-TOFMS as usual. An example of the G-reaction is shown in Figure 45 (SEQ ID Nos. 167-220) and the expected sequencing ladder is shown in Table IV with theoretical mass values for each ladder component. All sequencing peaks were well resolved except the last component (theoretical value 20577.4) was indistinguishable from the

-119-

background. Two neighboring sequencing peaks (a 62-mer and a 63-mer) were also separated indicating that such sequencing analysis could be applicable to longer templates. Again, an addition of an extra nucleotide by the Sequenase enzyme was observed in this spectrum.

- 5 This addition is not template specific and appeared in all four reactions which makes it easy to be identified. Compared to the primer peak, the sequencing peaks were at much lower intensity in the long template case.

TABLE IV

AAGATCTGACCAGGGATTCCGTTAGCGTGACTGCTGCTGCTGCTGATGATCCGACGCATCAGATCTGG-(A) <sub>n</sub> -3'	
1	3'-CTACTAGGCTGCGTAGTC-5'
2	3'-CCTACTAGGCTGCGTAGTC-5'
3	3'-ACCTACTAGGCTGCGTAGTC-5'
4	3'-GACCTACTAGGCTGCGTAGTC-5'
5	3'-CGACCTACTAGGCTGCGTAGTC-5'
6	3'-ACGACCTACTAGGCTGCGTAGTC-5'
7	3'-GACGACCTACTAGGCTGCGTAGTC-5'
8	3'-CGACGACCTACTAGGCTGCGTAGTC-5'
9	3'-ACGACGACCTACTAGGCTGCGTAGTC-5'
10	3'-GACGACGACCTACTAGGCTGCGTAGTC-5'
11	3'-CGACGACGACCTACTAGGCTGCGTAGTC-5'
12	3'-ACGACGACGACCTACTAGGCTGCGTAGTC-5'
13	3'-GACGACGACGACCTACTAGGCTGCGTAGTC-5'
14	3'-CGACGACGACGACCTACTAGGCTGCGTAGTC-5'

15	3'-ACGACGACGACGACCTACTAGGCTGCGTAGTC-5'
16	3'-GACGACGACGACGACCTACTAGGCTGCGTAGTC-5'
17	3'-CGACGACGACGACGACCTACTAGGCTGCGTAGTC-5'
18	3'-ACGACGACGACGACGACCTACTAGGCTGCGTAGTC-5'
19	3'-GACGACGACGACGACGACCTACTAGGCTGCGTAGTC-5'
20	3'-CGACGACGACGACGACGACCTACTAGGCTGCGTAGTC-5'
21	3'-ACGACGACGACGACGACGACCTACTAGGCTGCGTAGTC-5'
22	3'-GACGACGACGACGACGACGACCTACTAGGCTGCGTAGTC-5'
23	3'-CGACGACGACGACGACGACGACCTACTAGGCTGCGTAGTC-5'
24	3'-ACGACGACGACGACGACGACGACCTACTAGGCTGCGTAGTC-5'
25	3'-GACGACGACGACGACGACGACGACCTACTAGGCTGCGTAGTC-5'
26	3'-TGAACGACGACGACGACGACGACCTACTAGGCTGCGTAGTC-5'
27	3'-CTGACGACGACGACGACGACGACGACCTACTAGGCTGCGTAGTC-5'
28	3'-ACTGACGACGACGACGACGACGACGACCTACTAGGCTGCGTAGTC-5'
29	3'-CACTGACGACGACGACGACGACGACGACCTACTAGGCTGCGTAGTC-5'
30	3'-GCACTGACGACGACGACGACGACGACGACCTACTAGGCTGCGTAGTC-5'

5

10

15

31	3'-CGCACTGACGACGACGACGACGACGACGACTAGGTCGGTAGTC-5'
32	3'-TGCACGTGACGACGACGACGACGACGACGACTAGGTCGGTAGTC-5'
33	3'-ATCGCACTGACGACGACGACGACGACGACGACTAGGTCGGTAGTC-5'
34	3'-AATCGCACTGACGACGACGACGACGACGACGACTAGGTCGGTAGTC-5'
35	3'-CAATCGCACTGACGACGACGACGACGACGACGACTAGGTCGGTAGTC-5'
36	3'-CCAATCGCACTGACGACGACGACGACGACGACGACTAGGTCGGTAGTC-5'
37	3'-GCCAAATCGCACTGACGACGACGACGACGACGACGACTAGGTCGGTAGTC-5'
38	3'-AGCCAAATCGCACTGACGACGACGACGACGACGACGACTAGGTCGGTAGTC-5'
39	3'-AAGCCAATCGCACTGACGACGACGACGACGACGACGACTAGGTCGGTAGTC-5'
40	3'-TAAGCCAATCGCACTGACGACGACGACGACGACGACGACTAGGTCGGTAGTC-5'
41	3'-CTAAGCCAATCGCACTGACGACGACGACGACGACGACGACTAGGTCGGTAGTC-5'
42	3'-CCTAAGCCAATCGCACTGACGACGACGACGACGACGACGACTAGGTCGGTAGTC-5'
43	3'-CCCTAAGCCAATCGCACTGACGACGACGACGACGACGACGACTAGGTCGGTAGTC-5'
44	3'-TCCCTAAGCCAATCGCACTGACGACGACGACGACGACGACGACTAGGTCGGTAGTC-5'
45	3'-GTCCTAAGCCAATCGCACTGACGACGACGACGACGACGACGACTAGGTCGGTAGTC-5'
46	3'-GGTCCCTAAGCCAATCGCACTGACGACGACGACGACGACGACGACTAGGTCGGTAGTC-5'
47	3'-TGGTCCCTAAGCCAATCGCACTGACGACGACGACGACGACGACGACTAGGTCGGTAGTC-5'
48	3'-CTGGTCCCTAAGCCAATCGCACTGACGACGACGACGACGACGACGACTAGGTCGGTAGTC-5'
49	3'-ACTGGTCCCTAAGCCAATCGCACTGACGACGACGACGACGACGACGACTAGGTCGGTAGTC-5'
50	3'-GA TGGTCCCTAAGCCAATCGCACTGACGACGACGACGACGACGACGACTAGGTCGGTAGTC-5'
51	3'-AGACTGGTCCCTAAGCCAATCGCACTGACGACGACGACGACGACGACGACTAGGTCGGTAGTC-5'
52	3'-TAGACTGGTCCCTAAGCCAATCGCACTGACGACGACGACGACGACGACGACTAGGTCGGTAGTC-5'





-124-

TABLE IV Continued

	ddATP	ddCTP	ddGTP	ddTTP
1.	5491.6	5491.6	5491.6	5491.6
2.		5764.8		
5 3.	6078.0			
4.			6407.2	
5.		6696.4		
6.	7009.6			
7.			7338.8	
10 8.		7628.0		
9.	7941.2			
10.			8270.4	
11.		8559.6		
12.	8872.8			
15 13.			9202.0	
14.		9491.2		
15.	9804.4			
16.			10133.6	
17.		10422.88		
20 18.	10736.0			
19.			11065.2	
20.		11354.4		
21.	11667.6			
22.			11996.8	
25 23.		12286.0		
24.	12599.2			

-125-

5	25.			12928.4	
	26.				13232.6
	27.		13521.8		
	28.	13835.0			
	29.		14124.2		
10	30.			14453.4	
	31.		14742.6		
	32.				15046.8
	33.	15360.0			
	34.	15673.2			
15	35.		15962.4		
	36.		16251.6		
	37.			16580.8	
	38.	16894.0			
	39.	17207.2			
20	40.				17511.4
	41.		17800.6		
	42.		18189.8		
	43.		18379.0		
	44.				18683.2
25	45.			19012.4	
	46.			19341.6	
	47.				19645.8
	48.		19935.0		
	49.	20248.2			
	50.			20577.4	

-126-

51.	20890.6			
52.				21194.4
53.		21484.0		
54.				21788.2
55.				22092.4

#### Sequencing using duplex DNA probes for capturing and priming

- Duplex DNA probes with single-stranded overhang have been demonstrated to be able to capture specific DNA templates and also serve as primers for solid-state sequencing. The scheme is shown in Figure 46. Stacking interactions between a duplex probe and a single-stranded template allow only a 5-base overhang to be sufficient for capturing. Based on this format, a 5' fluorescent-labeled 23-mer (5'-GAT GAT CCG ACG CAT CAC AGC TC-3') (SEQ ID No. 29) was annealed to a 3'-biotinylated 18-mer (5'-GTG ATG CGT CGG ATC ATC-3') (SEQ ID No. 30), leaving a 5-base overhang. A 15-mer template (5'-TCG GTT CCA AGA GCT-3') (SEQ ID No. 31) was captured by the duplex and sequencing reactions were performed by extension of the 5-base overhang. MALDI-TOF mass spectra of the reactions are shown in Figure 47A-D. All sequencing peaks were resolved although at relatively low intensities. The last peak in each reaction is due to unspecific addition of one nucleotide to the full length extension product by the Sequenase enzyme. For comparison, the same products were run on a conventional DNA sequencer and a stacking fluorogram of the results is shown in Figure 48. As can be seen from the Figure, the mass spectra had the same pattern as the fluorogram with sequencing peaks at much lower intensity compared to the 23-mer primer.

-127-

**EXAMPLE 10****Thermo Sequenase Cycle Sequencing****MATERIALS AND METHODS**

*PCR amplification.* Human leukocytic genomic DNA was used for  
5 -PCR amplification. PCR primers to amplify a 209 bp fragment of the  
β-globin gene were the β2 forward primer (5'-CAT TTG CTT CTG ACA  
CAA CTG-3' SEQ ID NO. 32) and the β11 reverse primer (5'-CTT CTC  
TGT CTC CAC ATG C-3' SEQ ID NO. 33). Taq polymerase and 10x  
buffer were purchased from Boehringer-Mannheim (Germany) and dNTPs  
10 from Pharmacia (Freiburg, Germany). The total reaction volume was 50  
μl including 8 pmol of each primer with approximately 200 ng of genomic  
DNA used as template and a final dNTP concentration of 200 μM. PCR  
conditions were: 5 min at 94°C, followed by 40 cycles of 30 sec at  
94°C, 45 sec at 53°C, 30 sec at 72°C, and a final extension time of 2  
15 min at 72°C. The generated amplified product was purified and  
concentrated (2x) with the Qiagen 'Qiaquick' PCR purification kit  
(#28106) and stored in H<sub>2</sub>O.

*Cycle Sequencing.* Sequencing ladders were generated by  
primer extension with Thermo Sequenase™ -DNA Polymerase (Amersham  
20 LIFE Science, #E79000Y) under the following conditions: 7 pmol of HPLC  
purified primer (Cod5 12mer: 5'-TGC ACC TGA CTC-3' SEQ ID No. 34)  
were added to 6 μl purified and concentrated amplified product (i.e. 12 μl  
of the original amplified product), 2.5 units Thermo Sequenase and 2.5  
ml Thermo Sequenase reaction buffer in a total volume of 25 μl. The final  
25 nucleotide concentrations were 30 μM of the appropriate ddNTP (ddATP,  
ddCTP, ddGTP or ddTTP; Pharmacia Biotech, #27-2045-01) and 210 μM  
of each dNTP (7-deaza-dATP, DCTP, 7-deaza-GTP, dTTP; Pharmacia  
Biotech).

-128-

Cycling conditions were: denaturation for 4 min at 94°C, followed by 35 cycles of 30 sec at 94°C, 30 sec at 38°C, 30 sec at 55°C, and a final extension of 2 min at 72°C.

- Sample preparation and analysis by MALDI-TOF MS. After
- 5 completion of the cycling program, the reaction volume was increased to 50  $\mu$ l by addition of 25  $\mu$ l H<sub>2</sub>O. Desalting was achieved by shaking 30  $\mu$ l of ammonium saturated DOWEX (Fluka #44485) cation exchange beads with 50  $\mu$ l of the analyte for 2 min at room temperature. The Dowex beads, purchased in the protonated form, were pre-treated with 2M
- 10 NH<sub>4</sub>OH to convert them to the ammonium form, then washed with H<sub>2</sub>O until the supernatant was neutral, and finally put in 10 mM ammonium citrate for usage. After the cation exchange, DNA was purified and concentrated by ethanol precipitation by adding 5  $\mu$ l 3 M ammonium acetate (pH 6.5), 0.5  $\mu$ l glycogen (10 mg/ml, Sigma), and 110  $\mu$ l
- 15 absolute ethanol to the analyte and incubated at room temperature for 1 hour. After 12 min centrifugation at 20,000 X g the pellet was washed in 70% ethanol and resuspended in 1  $\mu$ l 18 Mohm/cm H<sub>2</sub>O water.

- For MALDI-TOF MS analysis 0.35  $\mu$ l of resuspended DNA was mixed with 0.35-1.3  $\mu$ l matrix solution (0.7 M 3-hydroxypicolinic acid
- 20 (3-HPA), 0.07 M ammonium citrate in 1:1 H<sub>2</sub>O:CH<sub>3</sub>CN) on a stainless steel sample target disk and allowed to air dry preceding spectrum acquisition using a Thermo Bioanalysis Vision 2000 MALDI-TOF operated in reflectron mode with 5 and 20 kV on the target and conversion dynode, respectively. External calibration generated from eight peaks
- 25 (3000-18000 Da) was used for all spectra.

## RESULTS

FIGURE 49 shows a MALDI-TOF mass spectrum of the sequencing ladder generated from a biological amplified product as template and a 12mer (5'-TGC ACC TGA CTC-3'(SEQ ID NO.34)) sequencing primer.

-129-

The peaks resulting from depurinations and peaks which are not related to the sequence are marked by an asterisk. MALDI-TOF MS measurements were taken on a reflectron TOF MS. A.) Sequencing ladder stopped with ddATP; B.) Sequencing ladder stopped with ddCTP; C.) Sequencing ladder stopped with ddGTP; D.) Sequencing ladder stopped with ddTTP.

FIGURE 50 shows a schematic representation of the sequencing ladder generated in Fig. 49 with the corresponding calculated molecular masses up to 40 bases after the primer (SEQ ID Nos 221-260). For the calculation the following masses were used: 3581.4 Da for the primer, 312.2 Da for 7-deaza-dATP, 304.2 Da for dTTP, 289.2 Da for dCTP and 328.2 Da for 7-deaza-dGTP.

FIGURE 51 shows the sequence of the amplified 209bp amplified product within the  $\beta$ -globin gene (SEQ ID No. 261), which was used as a template for sequencing. The sequences of the appropriate PCR primer and the location of the 12mer sequencing primer is also shown. This sequence represents a homozygote mutant at the position 4 after the primer. In a wildtype sequence this T would be replaced by an A.

#### EXAMPLE 11

#### 20 Microsatellite Analysis Using Primer Oligo Base Extension (PROBE) and MALDI-TOF Mass Spectrometry

##### SUMMARY

The method uses a single detection primer followed by an oligonucleotide extension step to give products differing in length by a number of bases specific for the number of repeat units or for second site mutations within the repeated region, which can be easily resolved by MALDI-TOF mass spectrometry. The method is demonstrated using as a model system the AluVpA polymorphism in intron 5 of the interferon- $\alpha$  receptor gene located on human chromosome 21, and the

-130-

poly T tract of the splice acceptor site of intron 8 from the CFTR gene located on human chromosome 7.

#### MATERIALS AND METHODS

Genomic DNA was obtained from 18 unrelated individuals and one family including of a mother, father, and three children. The repeated region was evaluated conventionally by denaturing gel electrophoresis and results obtained were confirmed by standard Sanger sequencing.

The primers for PCR amplification (8 pmol each) were IFNAR-IVS5-5': (5'-TGC TTA CTT AAC CCA GTG TG-3' SEQ ID. NO.35) and IFNAR-IVS5-3'.2: (5'-CAC ACT ATG TAA TAC TAT GC-3' SEQ ID. NO.36) for a part of the intron 5 of the interferon- $\alpha$  receptor gene, and CFEx9-F:(5'-GAA AAT ATC TGA CAA ACT CAT C-3' SEQ ID. NO.37) (5'-biotinylated) and CFEx9-R:(5'-CAT GGA CAC CAA ATT AAG TTC-3' SEQ ID. NO.38) for CFTR exon 9 with flanking intron sequences of the CFTR gene. Taq-polymerase including 10x buffer were purchased from Boehringer-Mannheim and dNTPs were obtained from Pharmacia. The total reaction volume was 50  $\mu$ l. PCR conditions were 5 min at 94°C followed by 40 cycles of: 1 min at 94°C, 45 sec at 53°C, and 30 sec at 72°C, and a final extension time of 5 min at 72°C.

Amplification products were purified using Qiagen's PCR purification kit (No.28106) according to manufacturer's instructions. Purified products were eluted from the column in 50  $\mu$ l TE-buffer (10mM Tris-HCl, 1 mM EDTA, pH 7.5).

A) Primer oligo base extension reaction (thermo cycling method) CyclePROBE was performed with 5 pmol appropriate detection primer (IFN:5'-TGA GAC TCT GTC TC-3' SEQ ID. NO.39) in a total volume of 25  $\mu$ l including 1 pmol purified template, 2 units Thermosequenase (Amersham Life Science, Cat. #E79000Y) 2.5  $\mu$ l Thermosequ nase buffer, 25  $\mu$ mol of each deoxynucleotide



-131-

(7-deaza-dATP, dTTP, and in some experiments extra dCTP) and 100  $\mu$ mol of dideoxyguanine and in some experiments additional ddCTP.

Cycling conditions: initial denaturation 94°C for 5 min followed by 30 cycles with 44°C annealing temperature for 30 sec and 55°C extension temperature for 1 min.

*Primer oligo base extension reaction (isothermal method)*

10  $\mu$ l aliquots of the purified double-stranded amplified product (~3 pmol) were transferred to a streptavidin-coated microliter plate well (~16 pmol capacity per 50  $\mu$ l volume; No. 1645684

- 10 Boehringer-Mannheim), followed by addition of 10  $\mu$ l incubation buffer (80 mM sodium phosphate, 400 mM NaCl, 0.4% Tween 20, pH 7.5) and 30  $\mu$ l water. After incubation for 1 hour at room temperature, the wells were washed three times with 200  $\mu$ l washing buffer A (40 mM Tris, 1 mM EDTA, 50 mM NaCl, 0.1% Tween 20, pH 8.8) and incubated with
- 15 100  $\mu$ l of 50 mM NaOH for 3 min to denature the double-stranded DNA. Finally, the wells were washed three times with 200  $\mu$ l 70 mM ammonium citrate solution.

- The annealing of 100 pmol detection primer (CFpT: 5'-TTC CCC AAA TCC CTG-3' SEQ ID NO. 40) was performed in 50  $\mu$ l annealing
- 20 buffer (50 mM ammonium phosphate buffer, pH 7.0 and 100 mM ammonium chloride) at 65°C for 2 min, at 37°C for 10 min, and at room temperature for 10 min. The wells were washed three times with 200  $\mu$ l washing buffer B (40 mM Tris, 1 mM EDTA, 50 mM NH<sub>4</sub>Cl, 0.1% Tween 20, pH 8.8) and once in 200  $\mu$ l TE buffer. The extension reaction was
- 25 performed using some components of the DNA sequencing kit from USB (No. 70770) and dNTPs or ddNTPs from Pharmacia. Total reaction volume was 45  $\mu$ l, containing of 21  $\mu$ l water, 6  $\mu$ l Sequenase-buffer, 3  $\mu$ l 100 mM DTT solution, 50  $\mu$ mol of 7-deaza-dATP, 20  $\mu$ mol ddCTP, 5.5  $\mu$ l glycerol enzyme dilution buffer, 0.25  $\mu$ l Sequenase 2.0, and 0.25  $\mu$ l

-132-

pyrophosphatase. The reaction was pipetted on ice and incubated for 15 min at room temperature and for 5 min at 37°C. Finally, the wells were washed three times with 200  $\mu$ l washing buffer B.

The extended primer was denatured from the template strand by heating at 80°C for 10 min in 50  $\mu$ l of a 50 mM ammonium hydroxide solution.

For precipitation, 10  $\mu$ l 3 M  $\text{NH}_4$ -acetate (pH 6.5), 0.5  $\mu$ l glycogen (10 mg/ml water, Sigma, Cat.#G1765), and 110  $\mu$ l absolute ethanol were added to the supernatant and incubated for 1 hour at room temperature.

After centrifugation at 13.000 g for 10 min the pellet was washed in 70% ethanol and resuspended in 1  $\mu$ l 18 Mohm/cm  $\text{H}_2\text{O}$  water.

Sample preparation was performed by mixing 0.6  $\mu$ l of matrix solution (0.7 M 3-hydroxypicolinic acid, 0.07 M dibasic ammonium citrate in 1:1  $\text{H}_2\text{O}:\text{CH}_3\text{CN}$ ) with 0.3  $\mu$ l of resuspended DNA/glycogen pellet on a sample target and allowed to air dry. Up to 20 samples were spotted on a probe target disk for introduction into the source region of a Thermo Bioanalysis (formerly Finnigan) Visions 2000 MALDI-TOF operated in reflectron mode with 5 and 20 kV on the target and conversion dynode, respectively. Theoretical average molecular mass ( $M_r(\text{calc})$ ) were calculated from atomic compositions; reported experimental  $M_r$  ( $M_r(\text{exp})$ ) values are those of the singly-protonated form, determined using external calibration.

## RESULTS

The aim of the experiments was to develop a fast and reliable method for the exact determination of the number of repeat units in microsatellites or the length of a mononucleotide stretch including the potential to detect second site mutations within the polymorphic region. Therefore, a special kind of DNA sequencing (primer oligo base extension, PROBE) was combined with the evaluation of the resulting

-133-

products by matrix-assisted laser desorption ionization (MALDI) mass spectrometry (MS). The time-of-flight (TOF) reflectron arrangement was chosen as a possible mass measurement system. As an initial feasibility study, an examination was performed first on an AluVpA repeat

- 5 polymorphism located in intron 5 of the human interferon- $\alpha$  receptor gene (cyclePROBE reaction) and second on the poly T tract located in intron 8 of the human CFTR gene (isothermal PROBE reaction).

A schematic presentation of the cyclePROBE experiment for the AluVpA repeat polymorphism is given in Figure 52. The extension of the  
10 antisense strand (SEQ ID No. 262) was performed with the sense strand serving as the template. The detection primer is underlined. In a family study co-dominant segregation of the various alleles could be demonstrated by the electrophoretic procedure as well as by the cyclePROBE method followed by mass spec analysis (Figure 53). Those  
15 alleles of the mother and child 2, for which direct electrophoresis of the amplified product indicated one of the two copies to have 13 repeat units, were measured using cyclePROBE to have instead only 11 units using ddG as terminator. The replacement of ddG by ddC resulted in a further unexpected short allele with a molecular mass of approximately  
20 11650 in the DNA of the mother and child 2 (Figure 54). Sequence analysis verified this presence of two second site mutations in the allele with 13 repeat units. The first is a C to T transition in the third repeat unit and the second mutation is a T to G transversion in the ninth repeat unit. Examination of 28 unrelated individuals shows that the 13 unit  
25 allele is spliced into a normal allele and a truncated allele using cyclePROBE. Statistical evaluation shows that the polymorphism is in Hardy-Weinberg equilibrium for both methods, however, using cyclePROBE as detection method the polymorphism information content is increased to 0.734.

-134-

PROBE was also used as an isothermic method for the detection of the three common alleles at the intron 8 splice acceptor site of the CFTR gene (SEQ ID No. 263). Figure 55 shows a schematic presentation of the expected diagnostic products (SEQ ID Nos. 264-266) with the theoretical mass values. The reaction was also performed in the antisense direction.

Figure 56 demonstrates that all three common alleles (T5, T7, and T9, respectively) at this locus could be reliably disclosed by this method. Reference to Figure 56 indicates that mass accuracy and precision with the reflectron time of flight used in this study ranged from 0-0.4%, with a relative standard deviation of 0.13%. This corresponds to far better than single base accuracy for the up to <90-mer diagnostic products generated in the IFNAR system. Such high analytical sensitivity is sufficient to detect single or multiple insertion/deletion mutations within the repeat unit or its flanking regions, which would induce >1% mass shifts in a 90-mer. This is analogous to the Figure 56 polyT tract analysis. Other mutations (i.e. an A to T or a T to A mutation within the IFNAR gene A3T repeat) which do not cause premature product termination are not detectable using any dNTP/ddNTP combination with PROBE and low performance MS instrumentation; a 9 Da shift in a 90-mer corresponds to a 0.03% mass shift. Achieving the accuracy and precision required to detect such minor mass shifts has been demonstrated with higher performance instrumentation such as Fourier transform (FT)MS, for which single Da accuracy is obtained up to 100-mers. Further, tandem FTMS, in which a mass shifted fragment can be isolated within the instrument and dissociated to generate sequence specific fragments, has been demonstrated to locate point mutations to the base in comparably sized products. Thus the combination of PROBE with higher performance instrumentation will have an analytical

-135-

sensitivity which can be matched only by cumbersome full sequencing of the repeat region.

### EXAMPLE 12

#### 5 Improved Apolipoprotein E Genotyping Using Primer Oligo Base Extension (PROBE) and MALDI-TOF Mass Spectrometry

#### MATERIALS AND METHODS

##### *PCR amplification.*

Human leukocytic genomic DNA from 100 anonymous individuals from a previously published study (Braun, A *et al.*, (1992) Human Genet. 89:401-406) were screened for apolipoprotein E genotypes using  
10 conventional methods. PCR primers to amplify a portion of exon 4 of the apo E gene were delineated according to the published sequence (Das, HK *et al.*, (1985) J. Biol. Chem. 260:6240-6247) (forward primer, apoE-F: 5'-GGC ACG GCT GTC CAA GGA G-3' SEQ ID. NO.41; reverse,  
15 apoE-R: 5'-AGG CCG CGC TCG GCG CCC TC-3' SEQ ID. NO.42). Taq polymerase and 10x buffer were purchased from Boehringer-Mannheim (Germany) and dNTPs from Pharmacia (Freiburg, Germany). The total reaction volume was 50  $\mu$ L including 8 pmol of each primer and 10% DMSO (dimethylsulfoxide, Sigma) with approximately 200 ng of genomic  
20 DNA used as template. Solutions were heated to 80°C before the addition of 1U polymerase; PCR conditions were: 2 min at 94°C, followed by 40 cycles of 30 sec at 94°C, 45 sec at 63°C, 30 sec at 72°C, and a final extension time of 2 min at 72°C.

##### *Restriction enzyme digestion and polyacrylamide electrophoresis.*

25 CfoI and RsaI and reaction buffer L were purchased from Boehringer-Mannheim, and HhaI from Pharmacia (Freiburg, Germany). For CfoI alone and simultaneous CfoI/RsaI digestion, 20  $\mu$ L of amplified products were diluted with 15  $\mu$ L water and 4  $\mu$ L Boehringer-Mannheim buffer L; after addition of 10 units of appropriate restriction enzyme(s)

-136-

the samples were incubated for 60 min at 37°C. The procedure for simultaneous HhaI/RsaI digestion required first digestion by RsaI in buffer L for one hour followed by addition of NaCl (50 mM end concentration) and HhaI, and additional incubation for one hour. 20 pL of the restriction digest were analyzed on a 12% polyacrylamide gel as described elsewhere (Hixson (1990) J. Lipid Res. 31:545-548). Recognition sequences of RsaI and CfoI (HhaI) are GT/AC and GCG/C, respectively; masses of expected digestion fragments from the 252-mer amplified product with CfoI alone and the simultaneous double digest with CfoI (or HhaI) and RsaI are given in Table V.

#### *Thermo-PROBE.*

PCR amplification was performed as described above, but with products purified with the Qiagen' Qiaquick' kit to remove unincorporated primers. Multiplex Thermo-PROBE was performed with 35 µl amplified product and 8 pmol each of the codon 112 (5'-GCG GAC ATG GAG GAC GTG-3' SEQ ID. NO.43) and 158 (5'-GAT GCC GAT GAC CTG CAG AAG-3' SEQ ID. NO.44) detection primers in 20 µl including ~1 pmol purified biotinylated antisense template immobilized on streptavidin coated magnetic beads, 2.5 units Thermosequencase, 2 µl Thermosequencase buffer, 50 µM of each dNTP and 200 µM of ddXTP, with the base identity of N and X as described in the text. Cycling conditions were: denaturation (94°C, 30 sec) followed by 30 cycles at 94°C (10 min) and 60°C (45 sec).

#### *Sample preparation and analysis by MALDI-TOF MS.*

For precipitation (Stults *et al.*, (1991) Rapid Commun. Mass Spectrom. 5: 359-363) of both digests and PROBE products, 5 µl 3M ammonium acetate (pH 6.5), 0.5 µl glycogen (10 mg/ml, Sigma), and 110 µl absolute ethanol were added to 50 µl of the analyte solutions and stored for 1 hour at room temperature. After 10 min centrifugation at

-137-

13,000 X g the pellet was washed in 70% ethanol and resuspended in 1  $\mu$ l 18 Mohm/cm H<sub>2</sub>O water. Where noted in the text, additional desalting was achieved by shaking 10-20  $\mu$ L of ammonium saturated DOWEX (Fluka #44485) cation exchange beads in 40  $\mu$ L of analyte. The beads, purchased in the protonated form, were pre-treated with three 5 min spin-decant steps in 2M NH<sub>4</sub>OH, followed with H<sub>2</sub>O and 10 mM ammonium citrate.

0.35  $\mu$ L of resuspended DNA was mixed with 0.35-1.3  $\mu$ L matrix solutions (Wu *et al.* (1993) Rapid Commun. Mass Spectrom. 7:142-146)

10 0.7 M 3-hydroxypicolinic acid (3-HPA), 0.07 M ammonium citrate in 1:1 H<sub>2</sub>O:CH<sub>3</sub>CN) on a stainless steel sample target disk and allowed to air dry preceding spectrum acquisition using a Thermo Bioanalysis Vision 2000 MALDI-TOF operated in reflectron mode with 5 and 20 kV on the target and conversion dynode, respectively. Theoretical average molecular

15 masses ( $M_r(\text{calc})$ ) of the fragments were calculated from atomic compositions; the mass of a proton (1.08 Da) is subtracted from raw data values in reporting experimental molecular masses ( $M_r(\text{exp})$ ) as neutral basis. An external calibration generated from eight peaks (3000-18000 Da) was applied to all spectra.

## 20 RESULTS

### *Digestion with Cfol alone.*

The inset to Figure 57a shows a 12% polyacrylamide gel electrophoretic separation of an  $\epsilon 3/\epsilon 3$  genotype after digestion of the 252 bp apo E amplified product with Cfol. Comparison of the

25 electrophoretic bands with a molecular weight ladder shows the cutting pattern to be as mostly as expected (Table V) for the  $\epsilon 3/\epsilon 3$  genotype. Differences are that the faint band at approximately 25 bp is not expected, and the smallest fragments are not observed. The accompanying mass spectrum of precipitated digest products shows a

-138-

similar pattern, albeit at higher resolution. Comparison with Table V shows that the observed masses are consistent with those of single-stranded DNA; the combination of an acidic matrix environment (3-HPA, pK<sub>a</sub> 3) and the absorption of thermal energy via interactions with the 337 nm absorbing 3-HPA upon ionization is known to denature short stretches of dsDNA under normal MALDI conditions (Tang, K *et al.*, (1994) *Rapid Commun Mass Spectrom* 8:183-186).

The approximately 25-mers, unresolved with electrophoresis, are resolved by MS as three single stranded fragments; while the largest (7427 Da) of these may represent a doubly charged ion from the 14.8 kDa fragments ( $m = 14850$ ,  $z = 2$ ;  $m/z = 7425$ ), the 6715 and 7153 Da fragments could result from PCR artifacts or primer impurities; all three peaks are not observed when amplified products are purified with Qiagen purification kits prior to digestion. The Table V 8871 Da 29-mer sense strand 3'-terminal fragment is not observed; the species detected at 9186 Da is consistent with the addition of an extra base (9187 - 8871 = 316, consistent with A) by the Taq-polymerase during PCR amplification (Hu, G *et al.*, (1993) *DNA and Cell Biol* 12:763-770). The individual single strands of each double strand with <35 bases (11 kDa) are resolved as single peaks; the 48-base single strands ( $M_r(\text{calc})$  14845 and 14858), however, are observed as an unresolved single peak at 14850 Da. Separating these into single peaks would require a mass resolution ( $m/\Delta m$ , the ratio of the mass to the peak width at half height) of  $14850/13 = 1140$ , nearly an order of magnitude greater than what is routine with the standard reflectron time-of-flight instrumentation used in this study; resolving such small mass differences with high performance instrumentation such as Fourier transform MS, which provides up to three orders of magnitude higher resolution in this mass range, has been demonstrated. The 91-mer single strands ( $M_r(\text{calc})$  27849 and 28436)



-139-

are also not resolved, even though this requires a resolution of only  $< 50$ . The dramatic decrease in peak quality at higher masses is due to metastable fragmentation (i.e. depurination) resulting from excess internal energy absorbed during and subsequent to laser irradiation.

5     ***Simultaneous digestion with Cfol and RsaI.***

Figure 57b (inset) shows a 12% polyacrylamide gel electrophoresis separation of  $\epsilon 3/\epsilon 3$  double digest products, with bands consistent with dsDNA with 24, 31, 36, 48, and 55 base pairs, but not for the smaller fragments. Although more peaks are generated (Table V) than with Cfol  
10 alone, the corresponding mass spectrum is more easily interpreted and reproducible since all fragments contain  $< 60$  bases, a size range far more appropriate for MALDI-MS if reasonably accurate  $M_r$  values (e.g., 0.1%) are desired. For fragments in this mass range, the mass measuring accuracy using external calibration is  $-0.1\%$  (i.e.  $\leq +10$  Da at  
15 10 kDa). Significant depurination (indicated in Figure by asterisk) is observed for all peaks above 10 kDa, but even the largest peak at 17171 Da is clearly resolved from its depurination peak so that an accurate  $M_r$  can be measured. Although molar concentrations of digest products should be identical, some discrimination against those fragments with  
20  $\leq 11$  bases is observed, probably due to their loss in the ethanol/glycogen precipitation step. The quality of MS results from simultaneous digestion with Cfol (or HhaI) and RsaI is superior to those with Cfol (or HhaI) alone, since the smaller fragments generated are good for higher mass accuracy measurements, and with all genotypes there is  
25 no possibility for dimer peaks overlapping with high mass diagnostic peaks. Since digestion by RsaI/Cfol and RsaI/HhaI produce the same restriction fragments but the former may be performed as a simultaneous digest since their buffer requirements are the same, this enzyme mixture was used for all subsequent genotyping by restriction digest protocols.

-140-

Table V

Mass and Copy Number of Expected Restriction Digest Products

Table Va Cfol Digestion<sup>a</sup>

	(+) (-)	e2/e2				
		e2/e2	e2/e2	e2/e2	e2/e2	e2/e2
5	5781, 5999	--	--	1	--	1
	10752, 10921	--	1	1	2	2
	14845, 14858	--	1	1	2	2
	22102, 22440	--	--	1	--	1
	25575, 25763	2	1	1	--	--
10	27849, 28436	2	2	1	2	1

Table Vb. Cfol/Rsal Digestion<sup>b</sup>

	(+) (-)	e2/e2				
		e2/e2	e2/e3	e2/e4	e3/e3	e3/e4
	3428, 4025	--	1	1	2	2
15	5283, 5880	--	--	1	--	1
	5781, 5999	--	--	1	--	1
	11279, 11627	2	2	1	2	1
	14845, 14858	--	1	1	2	2
20	18269, 18848	2	2	1	--	--

<sup>a</sup>Cfol Invariant fragment masses: 1848, 2177, 2186, 2435, 4924, 5004, 5412, 5750, 8871, 9628 Da.

-141-

<sup>a</sup>CfoI/RsaI Invariant fragment masses: 1848, 2177, 2186, 2436, 4924, 5004, 5412, 5750, 6745, 7510, 8871, 9628, 16240, 17175 Da.

Table VI				
	ddT M <sub>r</sub> (Calc)	ddT M <sub>r</sub> (Exp)	ddC M <sub>r</sub> (Calc)	ddC M <sub>r</sub> (Exp)
5 e2/e2	<sup>a</sup> 5918, <sup>b</sup> 6768	_____	<sup>a</sup> 6536, <sup>b</sup> 7387	_____
e2/e3	<sup>a</sup> 5918, <sup>b</sup> 6768, <sup>b</sup> 7965	5919, 6769, 7967	<sup>a</sup> 6536, <sup>b</sup> 6753, <sup>b</sup> 7387	6542, 6752, 7393
e2/e4	<sup>a</sup> 5918, <sup>b</sup> 6768, <sup>b</sup> 7965, <sup>a</sup> 8970	_____	<sup>a</sup> 5903, <sup>b</sup> 6536, <sup>b</sup> 6753, <sup>a</sup> 7387	_____
e3/e3	<sup>a</sup> 5918, <sup>b</sup> 7965	5918, 7966	<sup>a</sup> 6536, <sup>b</sup> 6753	6542, 6756
e3/e4	<sup>a</sup> 5918, <sup>b</sup> 7965, <sup>a</sup> 8970	5914, 7959, 8965	<sup>a</sup> 5903, <sup>b</sup> 6536, <sup>b</sup> 6753	5898, 6533, 6747
10 e4/e4	<sup>b</sup> 7965, <sup>a</sup> 8970	7966, 8969	<sup>a</sup> 5903, <sup>b</sup> 6753	5900, 6752

<sup>a</sup>From codon 112 detection primer (unextended 5629.7 Da).

<sup>b</sup>From codon 158 detection primer (unextended 6480.3 Da).

Dashed lines: this genotype not available from the analyzed pool of 100 patients.

- 15 Figure 58a-c shows the ApoE  $\epsilon 3/\epsilon 3$  genotype after digestion with CfoI and a variety of precipitation schemes; equal volume aliquots of the same amplified product were used for each. The sample treated with a single precipitation (Figure 58a) from an ammonium acetate/ethanol/glycogen solution results in a mass spectrum characterized by broad peaks, especially at high mass. The masses for intense peaks at 5.4, 10.7, and 20 14.9 kDa are 26 Da (0.5%), 61 Da (0.6%), and 45 Da (0.3%) Da higher, respectively, than the expected values; the resolution (the ratio of a peak width at half its total intensity to the measured mass of the peak) for each of these is  $\sim 50$ , and decreases with increasing mass. Such

-142-

observations are consistent with a high level of nonvolatile cation adduction; for the 10.8 kDa fragment, the observed mass shift is consistent with a greater than unit ratio of adducted:nonadducted molecular ions.

- 5 MS peaks from a sample redissolved and precipitated a second time are far sharper (Figure 58b), with resolution values nearly double those of the corresponding Figure 58a peaks. Mass accuracy values are also considerably improved; each is within 0.07% of its respective calculated values, close to the independently determined instrumental
- 10 limits for DNA measurement using 3-HPA as a matrix. Single (not shown) and double (Figure 58C) precipitations with isopropyl alcohol (IPA) instead of ethanol result in resolution and mass accuracy values comparable to those for corresponding ethanol precipitations, but enhanced levels of dimerization are observed, again potentially confusing
- 15 measurements when such dimers overlap with higher mass "diagnostics" monomers present in the solution. EtOH/ammonium acetate precipitation with glycogen as a nucleation agent results in nearly quantitative recovery of fragments except for the 7-mers, serving as a simultaneous concentration and desalting step prior to MS detection. Precipitation
- 20 from the same EtOH/ammonium acetate solutions in the absence of glycogen results in far poorer recovery, especially at low mass.

The results indicate that to obtain accurate ( $M_r(\text{exp})$ ) values after either IPA and EtOH precipitations, a second precipitation is necessary to maintain high mass accuracy and resolution.

- 25 The ratio of matrix:digest product also affects spectral quality; severe suppression of higher mass fragments (not shown) observed with 1:1 volume matrix: digest product (redissolved in 1  $\mu$ L) is alleviated by using a 3 - 5 fold volume excess of matrix.

-143-

*Apo E genotyping by enzymatic digestion.* Codon 112 and 158 polymorphisms fall within CfoI (but not RsaI) recognition sequences. In the 252 bp amplified product studied here, invariant (i.e. cut in all genotypes) sites cause cuts after bases 31, 47, 138, 156, 239, and 246. The cutting site after base 66 is only present for  $\epsilon 4$ , while that after base 204 is present in  $\epsilon 3$  and  $\epsilon 4$ ; the  $\epsilon 2$  genotype is cut at neither of these sites. These differences in the restriction pattern can be demonstrated as variations in mass spectra. Figure 59 shows mass spectra from several ApoE genotypes available from a pool of 100 patients (Braun, A *et al.*, (1992) Hum. Genet. 89:401-406). Vertical dashed lines are drawn through those masses corresponding to the expected Table V diagnostic fragments; other labeled fragments are invariant. Referring to Table V, note that a fragment is only considered "invariant" if it is present in duplicate copies for a given allele; to satisfy this requirement, such a fragment must be generated in each of the  $\epsilon 2$ ,  $\epsilon 3$ , and  $\epsilon 4$  alleles.

The spectrum in Figure 59a contains all of the expected invariant fragments above 3 kDa, as well as diagnostic peaks at 3428 and 4021 (both weak), 11276 and 11627 (both intense), 14845, 18271, and 18865 Da. The spectrum in Figure 59b is nearly identical except that the pair of peaks at 18 kDa is not detected, and the relative peak intensities, most notably among the 11-18 kDa fragments, are different. The spectrum in Figure 59c also has no 18 kDa fragments, but instead has new low intensity peaks between 5-6 kDa. The intensity ratios for fragments above 9 kDa are similar to those of Figure 59b except for a relatively lower 11 kDa fragment pair. Figure 59d, which again contains the 5-6 kDa cluster of peaks, is the only spectrum with no 11 kDa fragments, and like the previous two also has no 18 kDa fragment.

-144-

Despite the myriad of peaks in each spectrum, each genotype can be identified by the presence and absence of only a few of the Table Vb diagnostic peaks. Due to the limited resolution of the MALDI-TOF instrumentation employed, the most difficult genotypes to differentiate are those based upon the presence or absence of the four diagnostic fragments between 5.2 and 6.0 kDa characteristic of the  $\epsilon 4$  allele, since these fragments nearly overlap with several invariant peaks. It has been found herein that the 5283 Da diagnostic fragment overlaps with a depurination peak from the 5412 Da invariant fragment, and the 5781 Da diagnostic peak is normally not completely resolved from the 5750 Da invariant fragment. Thus, distinguishing between an  $\epsilon 2/\epsilon 4$  and  $\epsilon 2/\epsilon 3$ , or between an  $\epsilon 3/\epsilon 4$  and an  $\epsilon 3/\epsilon 3$  allele, relies upon the presence or absence of the 5880 and 5999 Da fragments. Each of these is present in Figures 59c and 59d, but not in 59a or 59b.

The genotype of each of the patients in Figure 59 can be more rapidly identified by reference to the flowchart in Figure 60. Consider the Figure 59a spectrum. The intense pair of peaks at 11 kDa discounts the possibility of homozygous  $\epsilon 4$ , but does not differentiate between the other five genotypes. Likewise, the presence of the unresolved 14.8 kDa fragments is inconsistent with homozygous  $\epsilon 2$ , but leaves four possibilities ( $\epsilon 2/\epsilon 3$ ,  $\epsilon 2/\epsilon 4$ ,  $\epsilon 3/\epsilon 3$ ,  $\epsilon 3/\epsilon 4$ ). Of these only  $\epsilon 2/\epsilon 3$  and  $\epsilon 2/\epsilon 4$  are consistent with the 18 kDa peaks; the lack of peaks at 5283, 5879, 5779, and 5998, Da indicate that the Figure 59a sample is  $\epsilon 2/\epsilon 3$ . Using the same procedure, the Figures 59b-d genotypes can be identified as  $\epsilon 3/\epsilon 3$ ,  $\epsilon 3/\epsilon 4$ , and  $\epsilon 4/\epsilon 4$ , respectively. To date, all allele identifications by this method have been consistent with, and in many cases more easily interpreted than, those attained via conventional methods. The assignment can be further confirmed by assuring that fragment intensity ratios are consistent with the copy numbers of Table V. For instance,

-145-

the 14.8 kDa fragments are of lower intensity than those at 16- 17 kDa in Figure 59a, but the opposite is seen in Figures 59b-d. This is as expected, since in the latter three genotypes the 14.8 kDa fragments are present in duplicate, but the first is a heterozygote containing  $\epsilon 2$ , so that  
5 half of the amplified products do not contribute to the 14.8 kDa signal. Likewise, comparison of the 11 kDa fragment intensity to those at 9.6 and 14.8 kDa indicate that this fragment is double, double, single, and zero copy in Figures 59a, d, respectively. These data confirm that MALDI can perform in a semi-quantitative way under these conditions.

10 *ApoE genotyping by Primer Oligo Base Extension (PROBE).* The PROBE reaction was also tested as a means of simultaneous detection of the codon 112 and 158 polymorphisms. A detection primer is annealed to a single-stranded PCR-amplified template so that its 3' terminus is just downstream of the variable site. Extension of this primer by a DNA  
15 polymerase in the presence of three dNTPs and one ddXTP (that is not present as a dNTP) results in products whose length and mass depend upon the identity of the polymorphic base. Unlike standard Sanger type sequencing, in which a particular base-specific tube contains -99% dXTP and -1% ddXTP, the PROBE mixture contains 100% of a particular  
20 ddXTP combined with the other three dNTPs. Thus with PROBE a full stop of all detection primers is achieved after the first base complementary to the ddXTP is reached.

For the  $\epsilon 2/\epsilon 3$  genotype, the PROBE reaction (mixture of ddTTP, dATP, dCTP, dGTP) causes a  $M_r(\text{exp})$  shift of the codon 112 primer to  
25 5919 Da, and of the codon 158 primer to 6769 and 7967 Da (Table VI); a pair of extension products results from the single codon 158 primer because the  $\epsilon 2/\epsilon 3$  genotype is heterozygous at this position. Three extension products (one from codon 158, two from 112) are also observed from the heterozygote  $\epsilon 3/\epsilon 4$  (Figure 61c and Table VI), while

-146-

only two products (one from each primer) are observed from the Figure 61b ( $\epsilon 3/\epsilon 3$ ) and Figure 59d ( $\epsilon 4/\epsilon 4$ ) homozygote alleles. Referring to Table VI, each of the available alleles result in all expected ddT reaction product masses within 0.1% of the theoretical mass, and thus each is  
5 unambiguously characterized by this data alone. Further configuration of the allele identities may be obtained by repeating the reaction with ddCTP (plus dATP, dTTP, dGTP); these results, summarized also in Table VI, unambiguously confirm the ddT results.

*Appropriateness of the methods.* Comparison of Figures 59  
10 (restriction digestion) and 61 (PROBE) indicates that the PROBE method provides far more easily interpreted spectra for the multiplex analysis of codon 112 and 158 polymorphisms than does the restriction digest analysis. While the digests generate up to ~25 peaks per mass spectrum and in some case diagnostic fragments overlapping with invariant  
15 fragments, the PROBE reaction generates a maximum of only two peaks per detection primer (i.e. polymorphism). Automated peak detection, spectrum analysis, and allele identification would clearly be far more straightforward for the latter. Spectra for highly multiplexed PROBE, in which several polymorphic sites from the same or different amplified  
20 products are measured from one tube, are also potentially simple to analyze. Underscoring its flexibility, PROBE data analysis can be further simplified by judicious a priori choice of primer lengths, which can be designed so that no primers or products can overlap in mass.

Thus while PROBE is the method of choice for large scale clinical  
25 testing of previously well characterized polymorphic sites, the restriction digest analysis as described here is ideally suited to screening for new mutations. The identity of each of the two polymorphisms discussed in this study affects the fragment pattern; if this is the only information used, then the MS detection is a faster alternative to conventional



-147-

electrophoretic separation of restriction fragment length polymorphism products. The exact measurement of fragment  $M_r$  values can also give information on about sites completely remote from the enzyme recognition site since other single point mutations necessarily alter the

5 mass of each of the single strands of the double stranded fragment containing the mutation. The 252 bp amplified product could also contain allelic variants resulting in, for example, previously described Gly127 Asp (Weisgraber, KH *et al.*, (1984) J. Clin. Invest. 73:1024-1033), Arg136Ser (Wardell, MR *et al.*, (1987) J. Clin. Invest. 80:483-10 490), Arg142Cys (Horie, Y *et al.*, (1992) J. Biol. Chem. 267:1962-1968), Arg145Cys (Rall SC Jr *et al.*, (1982) Proc. Natl. Acad. Sci. U.S.A. 79:4696-4700), Lys146Glu (Mann, WA *et al.*, (1995) J. Clin. Invest. 96:1100-1107), or Lys146Gln (Smit, M *et al.*, (1990) J. Lipid Res. 31:45-53) substitutions. The G→A base substitution which codes for the

15 Gly127 Asp amino acid substitution would result in a -16 Da shift in the sense strand, and in a +15 Da (C→T) shift in the antisense strand, but not in a change in the restriction pattern. Such a minor change would be virtually invisible by electrophoresis; however, with accurate mass determination the substitution could be detected; the invariant 55-mer

20 fragment at 16240 (sense) and 17175 Da would shift to 16224 and 17190 Da, respectively. Obtaining the mass accuracy required to detect such minor mass shifts using current MALDI-TOF instrumentation, even with internal calibration, is not routine since minor unresolved adducts and/or poorly defined peaks limit the ability for accurate mass calling.

25 With high performance electrospray ionization Fourier transform (ESI-FTMS) single Da accuracy has been achieved with synthetic oligonucleotides (Little, DP *et al.*, (1995) Proc. Natl. Acad. Sci. U.S.A. 92:2318-2322) up to 100-mers (Little, DP *et al.*, (1994) J. Am. Chem. Soc. 116:4893-4897), and similar results have recently been achieved

-148-

with up to 25-mers using MALDI-FTMS (Li, Y *et al.*, (1996) Anal. Chem. 68:2090-2096).

### EXAMPLE 13

#### 5 A Method for Mass Spectrometric Detection of DNA Fragments Associated With Telomerase Activity

##### INTRODUCTION

One-fourth of all deaths in the United States are due to malignant tumors (R.K. Jain, (1996) Science 271:1079-1080). For diagnostic and therapeutic purposes there is a high interest in reliable and sensitive  
10 methods of tumor cell detection.

Malignant cells can be distinguished from normal cells by different properties. One of those is the immortalization of malignant cells which enables uncontrolled cell-proliferation. Normal diploid mammalian cells undergo a finite number of population doublings in culture, before they  
15 undergo senescence. It is supposed that the number of population doublings in culture, before they undergo senescence. It is supposed that the number of population doublings is related to the shortening of chromosome ends, called telomers, in every cell division. The reason for said shortening is based on the properties of the conventional  
20 semiconservative replication machinery. DNA polymerases only work in 5' to 3' direction and need an RNA primer.

Immortalization is thought to be associated with the expression of active telomerase. Said telomerase is a ribonucleoprotein catalyzing repetitive elongation of templates. This activity can be detected in a  
25 native protein extract of telomerase containing cells by a special PCR-system (N.W. Kim *et al.* (1994) Science 266:2011-2015) known as telomeric repeat amplification protocol (TRAP). The assay, as used herein, is based on the telomerase specific extension of a substrate primer (TS) and a subsequent amplification of the telomerase specific

-149-

extension products by a PCR step using a second primer (bioCX) complementary to the repeat structure. The characteristic ladder fragments of those assays are conventionally detected by the use of gel electrophoretic and labeling or staining systems. These methods can be replaced by MALDI-TOF mass spectrometry leading to faster accurate and automated detection.

## MATERIALS AND METHODS

### *Preparation of cells*

1 x 10<sup>6</sup> cultured telomerase-positive cells were pelleted, washed once with PBS (137 mM NaCl, 2.7 mM KCl, 4.3 mM Na<sub>2</sub>HPO<sub>4</sub>•7H<sub>2</sub>O, 1.4 mM KH<sub>2</sub>PO<sub>4</sub> in sterile DEPC water). The prepared cells may be stored at -75°C. Tissue samples have to be homogenized, according to procedures well known in the art, before extraction.

### *Telomerase extraction*

Pellet was resuspended in 200 µl CHAPS lysis buffer (10 mM Tris-HCl pH 7.5, 1 mM MgCl<sub>2</sub>, 1 mM EGTA, 0.1 mM benzamidine, 5 mM β-mercaptoethanol, 0.5% CHAPS, 10% glycerol) and incubated on ice for 30 min. The sample was centrifuged at 12,000 g for 30 min at 4°C. The supernatant was transferred into a fresh tube and stored at 75°C until use.

### *TRAP-assay*

2 µl of telomerase extract were added to a mixture of 10x TRAP buffer (200 mM Tris-HCl pH 8.3, 15 mM MgCl<sub>2</sub>, 630 mM KCl, 0.05% Tween 20, 10 mM EGTA) 50x dNTP-mix (2.5 mM each dATP, dTTP, dGTP, and dCTP), 10 pmol of TS primer and 50 pmol of bio CX primer in a final volume of 50 µl. The mixture was incubated at 30°C for 10 minutes and 5 min. at 94°C, 2 units of Taq Polymerase were added and a PCR was performed with 30 cycles of 94°C for 30 seconds, 50°C for 30 seconds and 72°C for 45 seconds.

-150-

*Purification of TRAP-assay products*

For every TRAP-assay to be purified, 50  $\mu$ l Streptavidin M-280 Dynabeads (10 mg/ml) were washed twice with 1x BW buffer (5 mM Tris-HCl, pH 7.5, 0.5 mM EDTA, 1 M NaCl). 50  $\mu$ l of 2x BW buffer were added to the PCR mix and the beads were resuspended in this mixture. The beads were incubated under gentle shaking for 15 min. at ambient temperature. The supernatant was removed and the beads were washed twice with 1x BW buffer. To the beads 50  $\mu$ l 25% ammonium hydroxide were added and incubated at 60°C for 10 min. The supernatant was saved, the procedure repeated, both supernatants were pooled and 300  $\mu$ l ethanol (100%) were added. After 30 min. the DNA was pelleted at 13,000 rpm for 12 min., the pellet was air-dried and resuspended in 600 nl ultrapure water.

*MALDI-TOF MS of TRAP-assay products*

300 nl sample were mixed with 500 nl of saturated matrix-solution (3-HPA:ammonium citrate = 10:1 molar ratio in 50% aqueous acetonitrile), dried at ambient temperature and introduced into the mass spectrometer (Vision 2000, Finigan MAT). All spectra were collected in reflector mode using external calibration.

*Sequences and masses*

bioCX: d(bio-CCC TTA CCC TTA CCC TTA CCC TAA SEQ ID NO. 45), mass: 7540 Da.

TS: d(AAT CCG TGC AGC AGA GTT SEQ ID NO.46), mass: 5523 Da.

Telomeric-repeat structure: (TTAGGG)<sub>n</sub>, mass of one repeat: 1909.2

*Amplification products:*

TS elongated by three telomeric repeats (first amplification product): 12452 Da. (N<sub>3</sub>)

TS elongated by four telomeric repeats: 14361 Da. (N<sub>4</sub>)

TS elongated by seven telomeric repeats: 20088 Da. (N<sub>7</sub>)

-151-

## RESULTS

Figure 62 depicts a section of a TRAP-assay MALDI-TOF mass spectrum. Assigned are the primers TS and bioCX at 5497 and 7537 Da, respectively (calculated 5523 and 7540 Da). The signal marked by 5- an asterisk represents n-1 primer product of chemical DNA synthesis. The first telomerase specific TRAP-assay product is assigned at 12775 Da. This product represents a 40-mer containing three telomeric repeats. Due to primer sequences this is the first expected amplification product of a positive TRAP-assay. The product is elongated by an additional 10 nucleotide due to extendase activity of Taq DNA polymerase (calculated non-extended product: 12452 Da, by A extended product: 12765 Da). The signal at 6389 Da represents the doubly charged ion of this product (calculated: 6387 Da). Figure 63 shows a section of higher masses of the same spectrum as depicted in figure 62, therefore the signal at 15 12775 Da is identical to that in figure 62. The TRAP-assay product containing seven telomeric repeats, representing a 64-mer also elongated by an additional nucleotide, is detected at 20322 Da (calculated: 20395 Da). The signals marked 1, 2, 3 and 4 cannot be base-line resolved. This region includes of: 1. signal of dimeric n-1 primer, 2. second TRAP- 20 assay amplification product, containing 4 telomeric repeats and therefore representing a 46-mer (calculated: 14341 Da/14674 Da for extendase elongated product) and 3. dimeric primer-ion and furthermore all their corresponding depurination signals. There is a gap observed between the signals of the second and fifth extension product. This signal gap 25 corresponds to the reduced band intensities observed in some cases for the third and fourth extension product in autoradiographic analysis of TRAP-assays (N.W. Kim *et al.* (1994) *Science* 266:2013).

The above-mentioned problems, caused by the dimeric primer and related signals, can be overcome using an ultrafiltration step employing a

-152-

molecular weight cut-off membrane for primer removal prior to MALDI-TOF-MS analysis. This will permit an unambiguous assignment of the second amplification product.

#### EXAMPLE 14

##### 5 A method for Detecting Neuroblastoma-Specific Nested RT-amplified products Via MALDI-TOF Mass Spectrometry

###### Introduction

Neuroblastoma is predominantly a tumor of early childhood with 66% of the cases presenting in children younger than 5 years of age.

- 10 The most common symptoms are those due to tumor mass, bone pain, or those caused by excessive catecholamine secretion. In rare cases, neuroblastoma can be identified prenatally (R.W. Jennings et al, (1993) J. Ped. Surgery 28:1168-1174). Approximately 70% of all patients with neuroblastoma have metastatic disease at diagnosis. The prognosis is
- 15 dependent on age at diagnosis, clinical stage and other parameters.

For diagnostic purposes there is a high interest in reliable and sensitive methods of tumor cell detection, e.g., in control of autologous bone marrow transplants or on-going therapy.

- Since catecholamine synthesis is a characteristic property of
- 20 neuroblastoma cells and bone marrow cells lack this activity (H. Naito et al., (1991) Eur. J. Cancer 27:762-765), neuroblastoma cells or metastasis in bone marrow can be identified by detection of human tyrosine 3-hydroxylase (E.C. 1.14.16.2, hTH) which catalyzes the first step in biosynthesis of catecholamines.

- 25 The expression of hTH can be detected via reverse transcription (RT) polymerase chain reaction (PCR) and the amplified product can be analyzed via MALDI-TOF mass spectrometry.

###### Materials and methods

###### *Cell- or tissue-treatment*

-153-

Cultures cells were pelleted (10 min. 8000 rpm) and washed twice with PBS (137 mM NaCl, 2.7 mM KCl, 4.3 mM  $\text{Na}_2\text{HPO}_4 \cdot 7\text{H}_2\text{O}$ , 1.4 mM  $\text{KH}_2\text{PO}_4$  in sterile PEPC water). The pellet was resuspended in 1 ml lysis/binding buffer (100 mM Tris-HCl, pH 8.0, 500 mM LiCl, 10 mM EDTA, 1% Li-dodecyl sulfate, 5 mM DTT) until the solution becomes viscose. Viscosity was reduced by DNA-shear step using a 1 ml syringe. The lysate may be stored in  $-75^\circ\text{C}$  or processed further directly. Solid tissues (e.g., patient samples) have to be homogenized before lysis.

*Preparation of magnetic Oligo-dT(25) beads*

10        100  $\mu\text{L}$  beads per  $1 \times 10^6$  cells were separated from the storage buffer and washed twice with 200  $\mu\text{L}$  lysis/binding buffer.

*Isolation of poly A<sup>+</sup> RNA*

The cell lysate was added to the prepared beads and incubated for 5 min. at ambient temperature. The beads were separated magnetically for 2-5 min. and washed twice with 0.5 ml LDS (10 mM Tris-HCl, pH 8.0, 0.15 M LiCl, 1 mM EDTA, 0.1% LDS).

*Solid-phase first-strand cDNA synthesis*

The poly A<sup>+</sup>RNA containing beads were resuspended in 20  $\mu\text{L}$  of reverse transcription mix (50 mM Tris-HCl, pH 8.3, 8 mM  $\text{MgCl}_2$ , 30 mM KCl, 10 mM DTT, 1.7 mM dNTPs, 3 U AMV reverse transcriptase) and incubated for 1 hour at  $45^\circ\text{C}$  (with a resuspension step all ten min.). The beads were separated from the reverse transcription mix, resuspended in 50  $\mu\text{L}$  of elution buffer (2 mM EDTA pH 8.0) and heated to  $95^\circ\text{C}$  for 1 min. for elution of the RNA. The beads with the cDNA first-strand can be stored in TB (0.089 M Tris-base, 0.089 M boric acid, 0.2 mM EDTA pH 8.0), TE 10 mM Tris-HCl, 0.1 mM EDTA, pH 8.0) or 70% ethanol for further processing.

*Nested polymerase chain reaction*

-154-

Beads containing cDNA first-strand were washed twice with 1x PCR buffer (20 mM Tris-HCl pH 8.75, 10 mM KCl, 10 mM  $(\text{NH}_4)_2\text{SO}_4$ , 2 mM  $\text{MgSO}_4$ , 0.1% Triton X-100, 0.1 mg bovine serum albumin) and resuspended in PCR mix (containing 100) pmol of each outer primer, 2.5 u *Pfu* (exo-) DNA polymerase, 200  $\mu\text{M}$  of each dNTP and PCR buffer in a final volume of 50  $\mu\text{L}$ ). The mixture was incubated at 72°C 1 min. and amplified by PCR for 30 cycles. for the nested reaction: 1  $\mu\text{L}$  of the first PCR was added as template to a PCR mix d(as above but nested primers instead of outer primers) and subjected to the following temperature program: 94°C 1 min., 65°C 1 min. and 72°C 1 min. for 20 cycles.

*Purification of nested amplified products*

Primers and low-molecular reaction by-products are removed using 10,000 Da cut-off ultrafiltration-unit. Ultrafiltration was performed at 7,500 g for 25 minutes. For every PCR to be purified, 50  $\mu\text{L}$  Streptavidin M-280 Dynabeads (10 mg/ml) were washed twice with 1xBW buffer (5 mM Tris-HCl, pH 7.5, 0.5 mM EDTA, 1 M NaCl), added to the ultrafiltration membrane and incubated under gentle shaking for 15 min. at ambient temperature. The supernatant was removed and the beads were washed twice with 1xBW buffer. 50  $\mu\text{L}$  25% ammonium hydroxide were added to the beads and incubated at ambient temperature for 10 min. The supernatant was saved, the procedure repeated, both supernatants were pooled and 300  $\mu\text{L}$  ethanol (100%) were added. After 30 min. the DNA was pelleted at 13,000 rpm for 12 min., the pellet was air-dried and resuspended in 600 nl ultrapure water.

*MALDI-TOF MS of nested amplified products*

300 nl sample was mixed with 500 nl of saturated matrix-solution (3-HPA: ammonium citrate = 10:1 molar ratio in 50% aqueous acetonitrile), dried at ambient temperature and introduced into the mass



-155-

spectrometer (Vision 2000, Finigan MAT). All spectra were collected in reflector mode using external calibration.

Outer primers:

hTH1: d(TGT CAG AGC TGG ACA AGT GT SEQ ID NO:47)

5 hTH2: d(GAT ATT GTC TTC CCG GTA GC SEQ ID NO:48)

Nested primers:

bio-hTH d(bio-CTC GGA CCA GGT GTA CCG CC SEQ ID NO:49),  
mass: 6485 Da

10 hTH6; d(CCT GTA CTG GAA GGC GAT CTC SEQ ID NO:50),  
mass: 6422 21 Da

mass of biotinylated single strand amplified product: 19253.6 Da  
mass of nonbiotinylated single strand amplified product: 18758.2  
Da

Results

15 A MALDI-TOF mass spectrum of a human tyrosine 3-hydroxylase (hTH) specific nested amplified product (61-mer) is depicted in figure 64. The signal at 18763 Da corresponds to non-biotinylated strand of the amplified product (calculated: 18758.2 Da, mass error: 0.02 Da). The signals below 10,000 and above 35,000 Da are due to multiply charged  
20 and dimeric amplified product-ions, respectively.

The product was obtained from a solid phase cDNA derived in a reverse transcription reaction from  $1 \times 10^6$  cells of a neuroblastoma cell-line (L-A-N-1) as described above. The cDNA first-strand was subjected to a first PCR using outer primers (hTH1 and hTH2), an aliquot of this  
25 PCR was used as template in a second PCR using nested primers (biohTH and hTH6). The nested amplified product was purified and MALDI-TOF MS analyzed:

-156-

The spectrum in Fig. 64 demonstrates the possibility of neuroblastoma cell detection using nested RT-PCR and MALDI-TOF MS analysis.

#### EXAMPLE 15

##### 5. Rapid Detection of the RET Proto-oncogene Codon 634 Mutation Using Mass Spectrometry

###### Material and Methods

###### *Probe*

The identity of codon 634 in each of the three alleles was confirmed by RsaI enzymatic digestion, single strand conformational polymorphism or Sanger sequencing. Exon 11 of the RET gene was PCR amplified (40 cycles) from genomic DNA using Taq-Polymerase (Boehringer-Mannheim) with 8 pmol each of 5'-biotinylated forward (5'-biotin-CAT GAG GCA GAG CAT ACG CA-3' SEQ ID NO:51) and unmodified reverse (5'-GAC AGC AGC ACC GAG ACG AT-3' SEQ ID NO:52) primer per tube; amplified products were purified using the Qiagen (QIAquick" kit to remove unincorporated primers. 15  $\mu$ l of amplified product were immobilized on 10  $\mu$ L (10 mg/mL) Dynal streptavidin coated magnetic beads, denatured using the manufacturer's protocol, and the supernatant containing antisense strand discarded, the PROBE reaction was performed using thermoSequenase (TS) DNA Polymerase (Amersham) and Pharmacia dNTP/ddNTPs. 8 pmol of extension primer (5'-CGG CTG CGA TCA CCG TGC GG-3' SEQ ID NO:53) was added to 13  $\mu$ L H<sub>2</sub>O, 2  $\mu$ L TS-buffer, 2  $\mu$ L 2mM ddATP (or ddTTP), and 2  $\mu$ L of 0.5 mM dGTP/dCTP/dTTP (or dGTP/DCTP/dATP), and the mixture heated for 30 sec @ 94°C, followed by 30 cycles of 10 sec @ 94°C and 45 sec @ 50°C; after a 5 min. incubation @ 95°C, the supernatant was decanted, and products were desalted by ethanol precipitation with the addition of 0.5  $\mu$ L of 10mg/mL glycogen. The

-157-

resulting pellet was washed in 70% ethanol, air dried, and suspended in 1  $\mu$ L H<sub>2</sub>O. 300 nL of this was mixed with the MALDI matrix (0.7 M 3-hydroxypicolinic acid, 0.07 M ammonium citrate in 1:1 H<sub>2</sub>O:CH<sub>3</sub>CN) on a stainless steel sample probe and air dried. Mass spectra were collected on a Thermo Bionalysis Vision 2000 MALDI-TOF operated in reflectron mode with 5 and 20 kV on the target and conversion dynode, respectively. Experimental masses ( $m_{\text{exp}}$ ) reported are those of the neutral molecules as measured using external calibration.

*Direct Measurement of Diagnostic Products*

PCR amplifications conditions for a 44 bp region containing codon 634 were the same as above but using Pfu polymerase; the forward primer contained a ribonucleotide at its 3'-terminus (forward, 5'-GAT CCA CTG TGC GAC GAG C (SEQ ID NO:54) -ribo; reverse, 5'-GCG GCT GCG ATC ACC GTG C (SEQ ID NO:55). After product immobilization and washing, 80  $\mu$ L of 12.5% NH<sub>4</sub>OH was added and heated at 80°C overnight to cleave the primer from 44-mer (sense strand) to give a 25-mer. Supernatant was pipetted off while still hot, dried resuspended in 50  $\mu$ L H<sub>2</sub>O, precipitated, resuspended, and measured by MALDI-TOF as above. MALDI-FTMS spectra of 25-mer synthetic analogs were collected as previously described (Li, Y. *et al.*, (1996) *Anal. Chem.* 68:2090-2096); briefly, 1-10 pmol DNA was mixed 1:1 with matrix on a direct insertion probe, admitted into the external ion source (positive ion mode), ionized upon irradiance with a 337 nm wavelength laser pulse, and transferred via rf-only quadrupole rods into a 6.5 Tesla magnetic field where they were trapped collisionally. After a 15 second delay, ions were excited by a broadband chirp pulse and detected using 256K data points, resulting in time domain signals of 5 s duration. Reported (neutral) masses are those of the most abundant isotopic peak after subtracting the mass of the charge carrying proton (1.01 Da).

## Results

The first scheme presented utilizes the PROBE reaction shown schematically in Figure 65. A 20-mer primer is designed to bind specifically to a region on the complementary template downstream of the mutation site; upon annealing to the template, which is labelled with biotin and immobilized to streptavidin coated magnetic beads, the PROBE primer is presented with a mixture of the three deoxynucleotide triphosphates (dNTPs), a di-dNTP (ddNTP), and a DNA polymerase (Figure 65). The primer is extended by a series of bases specific to the identity of the variable base in codon 634; for any reaction mixture (e.g., ddA + dT + dC + dG), three possible extension products representing the three alleles are possible (Figure 65).

For the negative control (Figure 66), the PROBE reaction with ddATP + dNTPs (N = T, C, G) causes a  $M_r(\text{exp})$  shift of the primer from 6135 to 6726 Da ( $\Delta m = +591$ ). The absence of a peak at 6432 rules out a C→A mutation (Figure 65); the mass of the single observed peak is more consistent with extension by C-ddA ( $M_r(\text{calc})$  6721, +0.07% error) than by T-ddA ( $M_r(\text{calc})$  6736, -0.15% error) than of  $A_3TC_2G$  expected for C→A mutant. Combining the ddA and ddT reaction data, it is clear that the negative control is as expected homozygous normal at codon 634.

The ddA reaction for patient 1 also results in a single peak ( $M_r(\text{exp}) = 6731$ ) between expected values for wildtype and C→T mutation (Figure 65b). The ddT reaction, however, results in two clearly resolved peaks consistent with a heterozygote wildtype ( $M_r(\text{exp})$  8249, +0.04% mass error)/C→T mutant ( $M_r(\text{exp})$  6428 Da, +0.08% mass error). For patient 2, the pair of Figure 66c ddA products represent a heterozygote C→A ( $M_r(\text{exp})$  6431, -0.06% mass error)/normal ( $M_r(\text{exp})$  6719, -0.03% mass error) allele. The ddT reaction confirms this, with a

-159-

single peak measured at 8264 Da consistent with unresolved wildtype and C→A alleles. The value of duplicate experiments is seen by comparing Figures 66a and 66b; while for patient 1 the peak at 6726 from the ddA reaction represents only one species, similar peak from patient 1 is actually a pair of unresolved peaks differing in mass by 15 Da.

An alternate scheme for point mutation detection is differentiation of alleles by direct measurement of diagnostic product masses. A 44-mer containing the RET634 site was generated by the PCR, and the 19-mer sense primer removed by  $\text{NH}_4\text{OH}$  cleavage at a ribonucleotide at its 3' terminus.

Figure 67 shows a series of MALDI-FTMS spectra of synthetic analogs of short amplified products containing the RET634 mutant site. Figures 67a-c and 67d-f are homozygous and heterozygous genotypes, respectively. An internal calibration was done using the most abundant isotope peak for the wildtype allele; application of this (external) calibration to the five other spectra resulted in better than 20 ppm mass accuracy for each. Differentiation by mass alone of the alleles is straightforward, even for heterozygote mixtures whose components differ by 16.00 (Figure 67d), 2501 (Figure 67e), or 9.01 Da (Figure 65f). The value of high performance MS is clear when recognition of small DNA mass shifts is the basis for diagnosis of the presence or absence of a mutation. The recent reintroduction of delayed extraction (DE) techniques has improved the performance of MALDI-TOF with short DNAs (Roskey, M.T. *et al.*, (1996) *Anal. Chem.* 68:941-946); a resolving power (RP) of  $> 10^3$  has been reported for a mixed-base 50-mer, and a pair of 31-mers with a C or a T ( $\Delta m$  15 Da) at a variable position resolved nearly to baseline. Thus DE-TOF-MS has demonstrated the RP required for separation of the individual components of heterozygotes.

-160-

Even with DE, however, the precision of DNA mass measurement with TOF is typically 0.1% (8 Da at 8 kDa) using external calibration, sufficiently high to result in incorrect diagnoses. Despite the possibility of space charge induced frequency shifts (Marshall, A.G. *et al.* (1991) *Anal. Chem.* **63**:215A-229A), MALDI-FTMS mass errors are rarely as high as 0.005% (0.4 Da at 8 KDa), making internal calibration unnecessary.

The methods for DNA point mutation presented here are not only applicable to the analysis of single base mutations, but also to less demanding detection of single or multiple base insertions or deletions, and quantification of tandem two, three, or four base repeats. The PROBE reaction yields products amenable to analysis by relatively low performance ESI or MALDI instrumentation; direct measurement of short amplified product masses is an even more direct means of mutation detection, and will likely become more widespread with the increasing interest in high performance MS available with FTMS.

#### EXAMPLE 16

##### Immobilization of nucleic acids on solid supports via an acid-labile covalent bifunctional trityl linker

Aminolinked DNA was prepared and purified according to standard methods. A portion (10eq) was evaporated to dryness on a speedvac and suspended in anhydrous DMF/pyridine (9:1; 0.1 ml). To this was added the chlorotrityl chloride resin (1 eq, 1.05 $\mu$ mol/mg loading) and the mixture was shaken for 24 hours. The loading was checked by taking a sample of the resin, detritylating this using 80% AcOH, and measuring the absorbance at 260nm. Loading was ca. 150pmol/mg resin.

In 80% acetic acid, the half-life of cleavage was found to be substantially less than 5 minutes--this compares with trityl ether-based approaches of half-lives of 105 and 39 minutes for *para* and *meta*

-161-

substituted bifunctional dimethoxytrityl linkers respectively. Preliminary results have also indicated that the hydroxy picolinic acid matrix alone is sufficient to cleave the DNA from the chlorotrityl resin.

#### EXAMPLE 17

##### 5 Immobilization of nucleic acids on solid supports via hydrophobic trityl linker

The primer contained a 5'-dimethoxytrityl group attached using routine trityl-on DNA synthesis.

10 Cl8 beads from an oligo purification cartridge (0.2 mg) placed in a filter tip was washed with acetonitrile, then the solution of DNA (50 ng in 25  $\mu$ l) was flushed through. This was then washed with 5% acetonitrile in ammonium citrate buffer (70 mM, 250  $\mu$ l). To remove the DNA from the Cl8, the beads were washed with 40% acetonitrile in water (10 $\mu$ l) and concentrated to ca 2  $\mu$ l on the Speedvac. The sample  
15 was then submitted to MALDI.

The results showed that acetonitrile/water at levels of ca. > 30% are enough to dissociate the hydrophobic interaction. Since the matrix used in MALDI contains 50% acetonitrile, the DNA can be released from the support and successfully detected using MALDI-TOF MS (with the  
20 trityl group removed during the MALDI process).

Figure 69 is a schematic representation of nucleic acid immobilization via hydrophobic trityl linkers.

#### EXAMPLE 18

##### 25 Immobilization of nucleic acids on solid supports via Streptavidin-Iminobiotin

##### Experimental Procedure

2-iminobiotin N-hydroxy-succinimid ester (Sigma) was conjugated to the oligonucleotides with a 3'- or 5'-amino linker following the conditions suggested by the manufacturer. The completion of the

-162-

reaction was confirmed by MALDI-TOF MS analysis and the product was purified by reverse phase HPLC.

For each reaction, 0.1 mg of streptavidin-coated magnetic beads (Dynabeads M-280 Streptavidin from Dynal) were incubated with 80 pmol of the corresponding oligo in the presence of 1M NaCl and 50 mM ammonium carbonate (pH 9.5) at room temperature for one hour. The beads bound with oligonucleotides were washed twice with 50 mM ammonium carbonate (pH 9.5). Then the beads were incubated in 2  $\mu$ l of 3-HPA matrix at room temperature for 2 min. An aliquot of 0.5  $\mu$ l of supernatant was applied to MALDI-TOF. For biotin displacement experiment, 1.6. mol of free biotin (80-fold excess to the bound oligo) in 1  $\mu$ l of 50 mM ammonium citrate was added to the beads. After a 5 min. incubation at room temperature, 1  $\mu$ l of 3-HPA matrix was added and 0.5  $\mu$ l of supernatant was applied to MALDI-TOF MS. To maximize the recovery of the bound iminobiotin oligo, the beads from the above treatment were again incubated with a 2  $\mu$ l of 3-HPA matrix and 0.5  $\mu$ l of supernatant was applied to MALDI-TOF MS. The matrix alone and free biotin treatment quantitatively released iminobiotin oligo off the streptavidin beads as shown in Figures 70 and 71.

#### EXAMPLE 19

##### Mutation Analysis Using Loop Primer Oligo Base Extension

##### MATERIALS AND METHODS

**Genomic DNA.** Genomic DNA was obtained from healthy individuals and patients suffering from sickle cell anemia. The wildtype and mutated sequences have been evaluated conventionally by standard Sanger sequencing.

**PCR-Amplification.** PCR amplifications of a part of the  $\beta$ -globin was established and optimized to use the reaction product without a further purification step for capturing with streptavidin coated bead. The



-163-

- target amplification for LOOP-PROBE reactions were performed with the loop-cod5 d(GAG TCA GGT GCG CCA TGC CTC AAA CAG ACA CCA TGG CGC, SEQ ID No. 58) as forward primer and  $\beta$ -11-bio d(TCT CTG TCT CCA CAT GCC CAG, SEQ ID. No. 59) as biotinylated reverse
- 5- primer. The underlined nucleotide in the loop-cod5 primer is mutated to introduce an invariant CfoI restriction site into the amplicon and the nucleotides in italics are complementary to a part of the amplified product. The total PCR volume was 50  $\mu$ l including 200 ng genomic DNA, 1U Taq-polymerase (Boehringer-Mannheim, Cat# 1596594), 1.5
- 10 mM MgCl<sub>2</sub>, 0.2 mM dNTPs (Boehringer-Mannheim, Ca# 1277049), and 10 pmol of each primer. A specific fragment of the  $\beta$ -globin gene was amplified using the following cycling condition: 5 min 94°C followed by 40 cycles of : 30 sec @ 94°C, 30 sec @ 56°C, 30 sec @ 72°C, and a final extension of 2 min at 72°C.
- 15        *Capturing and denaturation of biotinylated templates.* 10  $\mu$ l paramagnetic beads coated with streptavidin (10mg/ml; Dynal, Dynabeads M-280 streptavidin Cat# 112.06) and treated with 5x binding solution (5 M NH<sub>4</sub>Cl, 0.3M NH<sub>4</sub>OH) were added to 40  $\mu$ l PCR volume (10  $\mu$ l of the amplified product was saved for check electrophoresis).
- 20 After incubation for 30 min at 37°C the supernatant was discarded. The captured templates were denatured with 50  $\mu$ l 100 mM NaOH for 5 min at ambient temperature, then washed once with 50  $\mu$ l 50 mM NH<sub>4</sub>OH and three times with 100  $\mu$ l 10mM Tris.Cl, pH 8.0. The single stranded DNA served as templates for PROBE reactions.
- 25        *Primer oligo base extension (PROBE) reaction.* The PROBE reactions were performed using Sequenase 2.0 (USB Cat# E70775Z including buffer) as enzyme and dNTPs and ddNTPs supplied by Boehringer-Mannheim (Cat# 1277049 and 1008382). The ratio between dNTPs (dCTP, dGTP, dTTP) and ddATP was 1:1 and the total used

-164-

concentration was 50  $\mu$ M of each nucleotide. After addition of 5  $\mu$ l 1-fold Sequenase-buffer the beads were incubated for 5 min at 65°C and for 10 min at 37°C. During this time the partially self complementary primer annealed with the target site. The enzymatic reaction started after  
5 addition of 0.5  $\mu$ l 100 mM dithiothreitol (DTT), 3.5  $\mu$ l dNTP/ddNTP solution, and 0.5  $\mu$ l Sequenase (0.8 U) and incubated at 37°C for 10 min. Hereafter, the beads were washed once in 1-fold TE buffer (10 mM Tris, 1mM EDTA, pH 8.0).

*CfoI restriction digest.* The restriction enzyme digest was  
10 performed in a total volume of 5  $\mu$ l using 10 U CfoI in 1-fold buffer L purchased from Boehringer-Mannheim. The incubation time was 20 min at 37°C.

*Conditioning of the diagnostic products for mass spectrometric analysis*

15 After the restriction digest, the supernatant was precipitated in 45  $\mu$ l H<sub>2</sub>O, 10  $\mu$ l 3M NH<sub>4</sub>-acetate (pH 6.5), 0.5  $\mu$ l glycogen (10 mg/ml in water, Sigma, Cat# G1765), and 110  $\mu$ l absolute ethanol for 1 hour at room temperature. After centrifugation at 13,000 g for 10 min the pellet was washed in 70% ethanol and resuspended in 2  $\mu$ l 18 Mohm/cm H<sub>2</sub>O.  
20 The beads were washed in 100  $\mu$ l 0.7 M NH<sub>4</sub> citrate followed by 100  $\mu$ l 0.05 M NH<sub>4</sub> citrate. The diagnostic products were obtained by heating the beads in 2  $\mu$ l 50 mM NH<sub>4</sub>OH at 80°C for 2 min.

*Sample preparation and analysis on MALDI-TOF mass spectrometry.*

25 Same preparation was performed by mixing 0.6  $\mu$ l of matrix solution (0.7 M 3-hydroxypicolinic acid, 0.07 M dibasic ammonium citrate in 1:1 H<sub>2</sub>O:CH<sub>3</sub>CN) with 0.3  $\mu$ l of either resuspended DNA/glycogen pellet or supernatant after heating the beads in 50 mM NH<sub>4</sub>OH on a sample target and allowed to air dry. The sample target

-165-

was automatically introduced in to the source region of an unmodified Perspective Voyager MALDI-TOF operated in delayed extraction linear mode with 5 and 20 kV on the target and conversion dynode, respectively. Theoretical molecular mass ( $M_r(\text{calc})$ ) were calculated from atomic compositions; reported experimental ( $M_r(\text{exp})$ ) values are those of the singly-protonated form.

## RESULTS

The LOOP-PROBE has been applied to the detection of the most common mutation of codon 6 of the human  $\beta$ -globin gene leading to sickle cell anemia. The single steps of the method are schematically presented in figure 72. For the analysis of codon 6, a part of the  $\beta$ -globin gene was amplified by PCR using the biotinylated reverse primer  $\beta 11\text{bio}$  and the primer loop-cod5 which is modified to introduce a *CfoI* recognition site (fig. 72a). The amplified product is 192 bp in length.

After PCR the amplification product was bound to streptavidin coated paramagnetic particles as described above. The antisense strand was isolated by denaturation of the double stranded amplified product (Fig. 72b). The intra-molecule annealing of the complementary 3' end was accomplished by a short heat denaturation step and incubation at 37°C.

The 3' end of the antisense strand is now partially double stranded (Fig. 72c). For analyzing the DNA downstream of the self annealed 3'-end of the antisense strand, the primer oligo base extension (PROBE) has been performed using ddATP, dCTP, dGTP, dTTP (Fig. 72d). This generates different products in length specific for the genotype of the analyzed individual. Before the determination of the length of these diagnostic products, the DNA was incubated with the *CfoI* restriction endonuclease that cuts 5' of the extended product. This step frees the stem loop from the template DNA whereas the extended product still keeps attached to

-166-

the template. The extended products are then denatured by heating from the template stand and analyzed by MALDI-TOF mass spectrometry.

Since the MALDI-TOF analyses were performed with a non-calibrated instrument, the mass deviation between observed and expected values was approximately 0.6% higher than theoretically calculated. Nevertheless, the results obtained were conclusive and reproducible within repeated experiments. In all analyzed supernatants after the restriction digest the stem loop could be detected. Independent of the genotype, the stem loop has had in all analyses molecular masses about 8150 Da (expected 8111 Da). An example is shown in Figure 73a. The second peak in this figure with a mass of 4076 Da is a doubly charged ion of the stem loop. Figure 73b to 73d show the analyses of different genotypes as indicated in the respective inserts. HbA is the wildtype genotype and HbC and HbS are two different mutations in codon 6 of the  $\beta$ -globin gene which cause sickle cell disease. In the wildtype situation a single peak with a molecular mass of 4247 Da and another with 6696 Da are detected (Fig. 73b). The latter corresponds to the biotinylated PCR primer ( $\beta$ -11-bio) unused in the PCR reaction which also has been removed in some experiments. The former corresponds to the diagnostic product for HbA. The analyses of the two individual DNA molecules with HbS trait as well as compound heterozygosity (HbS/HbC) for the sickle cell disorder lead also to unambiguous expected results (Fig. 73c and 73d).

In conclusion, the LOOP-PROBE is a powerful means for detection of mutations especially predominant disease causing mutations or common polymorphisms. The technique eliminates one specific reagent for mutation detection and, therefore, simplifies the process and makes it more amenable to automation. The specific extended product that is analyzed is cleaved off from the primer and is therefore shorter compared

-167-

to the conventional method. In addition, the annealing efficiency is higher compared to annealing of an added primer and should therefore generate more product. The process is compatible with multiplexing and various detection schemes (e.g., single base extension, oligo base extension and sequencing). For example, the extension of the loop-primer can be used for generation of short diagnostic sequencing ladders within highly polymorphic regions to perform, for example, HLA typing or resistance as well as species typing (e.g., *Mycobacterium tuberculosis*)).

#### EXAMPLE 20

#### 10 T7-RNA Polymerase Dependent Amplification of CKR-5 and Detection by MALDI-MS

##### MATERIALS AND METHODS

*Genomic DNA.* Human genomic DNA was obtained from healthy individuals.

15 *PCR-Amplification and Purification.* PCR amplification of a part of the CKR-5 gene was accomplished using ckrT7f as sense primer d(ACC TAG CGT TCA GTT CGA CTG AGA TAA TAC GAC TCA CTA TAG CAG CTC TCA TTT TCC ATA C (SEQ ID NO. 60). The underlined sequence corresponds to the sequence homologous to CKR-5, the bolded sequence corresponds to the T7-RNA polymerase promoter sequence and the italic sequence was chosen randomly. ckr5r was used as antisense primer d(AAC TAA GCC ATG TGC ACA ACA (SEQ ID NO. 61). Purification of the amplified product and removal of unincorporated nucleotides was carried out using the QIAquick purification kit (Qiagen, cat# 28104). In 20 the final PCR volume of 50  $\mu$ l were 200 ng genomic DNA, 1U Taq-polymerase (Boehringer-Mannheim, cat# 1596594), 1.5 mM MgCl<sub>2</sub> 0.2 mM dNTPs (Boehringer-Mannheim, cat# 1277049), and 10 pmol of each primer. The specific fragment of the CKR-5 gene was amplified using the 25 following cycling conditions: 5 min @ 94°C followed by 40 cycles of 45

-168-

sec @ 94°C, 45 sec 52°C, 5 sec @ 72°C, and a final extension of 5 min at 72°C.

*T7-RNA Polymerase conditions.* One third of the purified DNA (about 60ng) was used in the T7-RNA polymerase reaction. (Boehringer-Mannheim, cat# 881 767). The reaction was carried out for 2h at 37°C according to the manufacturer's conditions using the included buffer. The final reaction volume was 20  $\mu$ l. 0.7  $\mu$ l RNasin (33 U/ $\mu$ l) had been added. After the extension reaction, the enzyme was inactivated by incubation for 5 min at 65°C.

10      *DNA digestion and conditioning of the diagnostic products for mass spec analysis.*

The template DNA was digested by adding RNase-free DNase I (Boehringer-Mannheim, cat# 776 758) to the inactivated T7 mixture and incubation for 20 min at room temperature. Precipitation was carried out by adding 1  $\mu$ l glycogen (10 mg/ml, Sigma, cat# G1765), 1/10 volume 3M NH<sub>4</sub> acetate (pH 6.5), and 3 volume absolute ethanol and incubation for 1 hour at room temperature. After centrifugation at 13,000 g for 10 min, the pellet was washed in 70% ethanol and resuspended in 3  $\mu$ l 18 Mohm/cm H<sub>2</sub>O. 1  $\mu$ l was analyzed on an agarose gel.

20      *Sample preparation and analysis on MALDI-TOF mass spectrometry*

Sample preparation was performed by mixing 0.6  $\mu$ l of matrix solution (0.7 M 3-hydroxypicolinic acid, 0.07 M dibasic ammonium citrate in 1:1 H<sub>2</sub>O:CH<sub>3</sub>CN) with 0.3  $\mu$ l of resuspended DNA/glycogen on a sample target and allowed to air dry. The sample target was introduced into the source region of an unmodified Finnigan VISION2000 MALDI-TOF operated in relectron mode with 5kV. The theoretical molecular mass was calculated from atomic composition; reported experimental values are those of singly-protonated form.

30      **RESULTS**

-169-

The chemokine receptor CKR-5 has been identified as a major coreceptor in HIV-1 (see e.g., WO 96/39437 to Human Genome Sciences; Cohen, J. *et al.* Science 275:1261). A mutant allele that is characterized by a 32 bp deletion is found in 16% of the HIV-1 seronegative population whereas the frequency of this allele is 35% lower in the HIV-1 seropositive population. It is assumed that individuals homozygous for this allele are resistant to HIV-1. The T7-RNA polymerase dependent amplification was applied to identify this specific region of the chemokine receptor CKR-5 (Figure 74). Human genomic DNA was amplified using conventional PCR. The sense primer has been modified so that it contains a random sequence of 24 bases that facilitate polymerase binding and the T7-RNA polymerase promoter sequence (Figure 75). The putative start of transcription is at the first base 5' of the promoter sequence. ckr5r was used as an antisense primer. PCR conditions are outlined above. The amplified product derived from wildtype alleles is 75 bp in length. Primer and nucleotides were separated from the amplification product using the Qiagen QIAquick purification kit. One third of the purified product was applied to in vitro transcription with T7-RNA polymerase. To circumvent interference of the template DNA, it was digested by adding RNase-free DNase I. RNA was precipitated and this step also leaves the degraded DNA in the supernatant. Part of the redissolved RNA was analyzed on an agarose gel and the rest of the sample was prepared for MALDI-TOF analysis. The expected calculated mass of the product is 24560 Da. A dominant peak, that corresponds to an approximate mass of 25378.5 Da can be observed. Since the peak is very broad, an accurate determination of molecular mass was not possible. The peak does not correspond to residual DNA template. First, the template DNA is digested, and second,

-170-

the DNA strands would have a mass of 23036.0 and 23174 Da, respectively.

This example shows that T7 RNA polymerase can effectively amplify target DNA. The generated RNA can be detected by Mass spectrometry. In conjunction with modified (e.g., 3'-deoxy)ribonucleotides that are specifically incorporated by a RNA polymerase but not extended any further, this method can be applied to determine the sequence of a template DNA.

#### EXAMPLE 21

#### 10 MALDI Mass Spectrometry of RNA Endonuclease Digests

##### MATERIALS

Synthetic RNA (Sample A:5'-UCCGGUCUGAUGAGUCCGUGAGGAC-3' (SEQ ID 62); sample B:5'-GUCACUACAGGUGAGCUCCA-3' (SEQ ID NO 63); sample C:5'-CCAUGCGAGAGUAAGUAGUA-3' (SEQ ID NO. 64)) samples were obtained from DNA technology (Aarhus, Denmark) and purified on a denaturing polyacrylamide gel (Shaler, T. A. *et al.* (1996) Anal. Chem. 63:5766-579). RNases T<sub>1</sub> (Eurogentec), U<sub>2</sub> (Calbiochem), A (Boehringer-Mannheim) and PhyM (Pharmacia) were used without additional purification. Streptavidin-coated magnetic beads (Dynabeads M-280 Streptavidin, Dynal) were supplied as a suspension of 6-7 x 10<sup>8</sup> bead/ml (10 mg/ml) dissolved in phosphate-buffered saline (PBS) containing 0.1% BSA and 0.02% NaN<sub>3</sub>. 3-Hydroxypicolinic acid (3-HPA) (Aldrich) was purified by a separate desalting step before use as described in more detail elsewhere (Little, D. P. *et al.* (1995) Proc. Natl. Acad. Sci. U.S.A. 92, 2318-2322).



-171-

## METHODS

*In vitro* transcription reaction. The 5'-biotinylated 49 nt *in vitro* transcript (SEQ ID No. 65):

AGGCCUGCGGCAAGACGGAAAGACCAUGGUCCCUNAUCUGCCGCAGGAUC

- 5 was produced by transcription of the plasmid pUTMS2 (linearized with the restriction enzyme BamHI) with T7 RNA polymerase (Promega). For the transcription reaction 3  $\mu$ g template DNA and 50u T7 RNA polymerase were used in a 50  $\mu$ l volume of 1u/ $\mu$ l RNA guard (Rnax inhibitor, Pharmacia), 0.5 mM NTP's 1.0 mM 5'-biotin-ApG dinucleotide, 10 40 mM Tris-HCl (pH 8.0), 6 mM MgCl<sub>2</sub> 2 mM spermidine and 10 mM DTT. Incubation was performed at 37°C for 1 hour, then another aliquot of 50 units T7 RNA polymerase was added and incubation was continued for another hour. The mixture was adjusted to 2M NH<sub>4</sub> acetate and the RNA was precipitated by addition of one volume of ethanol and one 15 volume of isopropanol. The precipitated RNA was collected by centrifugation at 20,000 X g for 90 min at 4°C, the pellet was washed with 70% ethanol, dried and redissolved at 8 M urea. Further purification was achieved by electrophoresis through a denaturing polyacrylamide gel as described elsewhere (Shaler, T. A. *et al.* (1996) 20 Anal. Chem. 68:576-579). The ration of 5'-biotinylated to non-biotinylated transcripts was about 3:1.

- Ribonuclease assay.* For partial digestion with selected RNases different enzyme concentrations ad assay conditions were employed as summarized in table VII. The solvents for each enzyme were selected 25 following the suppliers' instructions. The concentrations of the synthetic RNA samples and the *in vitro* transcript were adjusted to 5-10 x 10<sup>-6</sup>M.

TABLE VII  
Overview and Assay Conditions of the RNases

Rnase	Source	Concentration [units Rnase/ $\mu$ gRNA]	Conditions	Incubation Time (max. number of fragments)	References
T <sub>1</sub>	<i>Aspergillus oryzae</i>	0.2	20 mM Tris-HCl, pH 5.7, 37°C	5 min.	Donis-Keller, H. et al., (1977) Nuc. Acids Res. 4:2527- 2537
U <sub>2</sub>	<i>Ustilago Sphaerogena</i>	0.01	20 mM DAC, pH 5.0, 37°C	15 min	Donis-Keller, H. et al., (1977) Nuc. Acids Res. 4:2527- 2537
PhyM	<i>Physarum polycephalum</i>	20	20 mM DAC, pH 5.0, 50°C	15 min	Donis-Keller, H. et al., (1980) Nuc. Acids Res. 8:3133- 3142
A	bovine pancreas	$4 \times 10^{-3}$	10 mM Tris-HCl, pH 7.5, 37°C	30 min	Breslow, R. and R. Xu. (1993) Proc. Natl. Acad. Sci. USA 90:1201-1207
CL <sub>3</sub>	chicken liver	0.01	10 mM Tris-HCl, pH 6.5, 37°C	30 min	Boguski, et al., (1980) J. Biol. Chem. 255:2160-2163
cusativin	<i>cucumis sativus</i> L.	0.05 ng	10 mM Tris-HCl, pH 6.5, 37°C	30 min	Rojas, M.A. et al. (1994) Planta 194:328-338

-173-

The reaction was stopped at selected times by mixing 0.6  $\mu$ l aliquots of the assay with 1.5  $\mu$ l of 3 HPA-solution. The solvent was subsequently evaporated in a stream of cold air for the MALDI-MS analysis.

- 5        Limited alkaline hydrolysis was performed by mixing equal volumes (2.0  $\mu$ l) of 25% ammonium hydroxide and RNA sample ( $5-10 \times 10^{-6}$  M) at 60°C. 1  $\mu$ l aliquots were taken out at selected times and dried in a stream of cold air. For these samples it turned out to be important to first dry the digests in a stream of cold air, before 1.5  $\mu$ l of the matrix solution and 0.7  $\mu$ l of  $\text{NH}_4^+$  loaded cation exchanged polymer beads were added.

- 15       The reaction was stopped at selected times by mixing 0.6  $\mu$ l aliquots of the assay with 1.5  $\mu$ l of 3HPA-solution. The solvent was subsequently evaporated in a stream of cold air for the MALDI-MS analysis.

- 20       Limited alkaline hydrolysis was performed by mixing equal volumes (2.0  $\mu$ l) of 25% ammonium hydroxide and RNA sample ( $5-10 \times 10^{-6}$  M) at 60°C. 1  $\mu$ l aliquots were taken out at selected times and dried in a stream of cold air. For these samples it turned out to be important to first dry the digests in a stream of cold air, before 1.5  $\mu$ l of the matrix solution and 0.7  $\mu$ l of a suspension of  $\text{NH}_4^+$  loaded cation exchange polymer beads were added.

- 25       *Separation of 5'-biotinylated fragments.* Streptavidin-coated magnetic beads were utilized to separate 5'-biotinylated fragments of the *in vitro* transcript after partial RNase degradation. The biotin moiety in

-174-

- this sample was introduced during the transcription reaction initiated by the 5'-biotin-pApG-dinucleotide. Prior to use, the beads were washed twice with 2 x binding & washing (b&w) buffer (20 mM Tris-HCl, 2 mM EDTA, 2 M NaCl pH 8.2) and resuspended at 10 mg/ml in 2 x b&w buffer. Circa 25 pmol of the RNA *in vitro* transcript were digested by RNase U2 using the protocol described above. The digestion was stopped by adding 3  $\mu$ l of 95% formamide containing 10 mM trans- 1,2-diaminocyclohexane-N,N,N',N'-tetraacetic acid (CDTA) at 90°C for 5 min, followed by cooling on ice. Subsequently, capture of the biotinylated fragments was achieved by incubation of 6  $\mu$ l of the digest with 6  $\mu$ l of the bead suspension and 3  $\mu$ l of b&w buffer at room temperature for 15 min. Given the binding capacity of the beads of 200 pmol of biotinylated oligonucleotide per mg of beads, as specified by the manufacturer, the almost 2-times excess of oligonucleotide was used to assure a full loading of the beads. The supernatant was removed, and the beads were washed twice with 6  $\mu$ l of H<sub>2</sub>O. The CDTA and 95% formamide at 90°C for 5 min. After evaporation of the solvent and the formamide the  $\leq 2.5$  pmol of fragments were resuspended in 2  $\mu$ l H<sub>2</sub>O and analyzed by MALDI-MS as described above.
- Sample preparation for MALDI-MS.* 3-Hydroxypicolinic acid (3-HPA) was dissolved in ultra pure water to a concentration of ca. 300 mM. Metal cations were exchanged against NH<sub>4</sub><sup>+</sup> as described in detail previously. (Little, D. P. *et al.* (1995)Proc. Natl. Acad. Sci. U.S.A. 92: 2318-2322). Aliquots of 0.6  $\mu$ l of the analyte solution were mixed with 1.5  $\mu$ l 3-HPA on a flat inert metal substrate. Remaining alkali cations,

-175-

present in the sample solution as well as on the substrate surface, were removed by the addition of 0.7  $\mu$ l of the solution of  $\text{NH}_4^+$  - loaded cation exchange polymer beads. During solvent evaporation, the beads accumulated in the center of the preparation, were not used for the  
5 analysis, and were easily removed with a pipette tip.

*Instrument.* A prototype of the Vision 2000 (ThermBioanalysis, Hemel, Hempstead, UK) reflectron time of flight mass spectrometer was used for the mass spectrometry. Ions were generated by irradiation with a frequency-tripled ND:YAG laser (355 nm, 5 ns; Spektrum GmbH,  
10 Berlin, Germany) and accelerated to 10 ke V. Delayed ion extraction was used for the acquisition of the spectra shown, as it was found to substantially enhance the signal to noise ratio and/or signal intensity. The equivalent flight path length of the system is 1.7 m, the base pressure is  $10^{-4}$  Pa. Ions were detected with a discrete dynode  
15 secondary-electron multiplier (R2362, Hamamatsu Photonics), equipped with a conversion dynode for effective detection of high mass ions. The total impact energy of the ions on the conversion dynode was adjusted to values ranging from 16 to 25 keV, depending on the mass to be detected. The preamplified output signal of the SEM was digitized by a  
20 LeCroy 9450 transient recorder (LeCroy, Chestnut Ridge, NY, USA) with a sampling rate of up to 400 MHz. For storage and further evaluation, the data were transferred to a personal computer equipped with custom-made software (ULISSES). All spectra shown were taken in the positive ion mode. Between 20 and 30 single shot spectra were averaged for  
25 each of the spectra shown.

-176-

## RESULTS

*Specificity of Rnases.* Combining base-specific RNA cleavage with MALDI-MS requires reaction conditions optimized to retain the activity and specificity of the selected enzymes on the one hand and complying with the boundary conditions for MALDI on the other. Incompatibility mainly results because the alkaline-ion buffers, commonly used in the described reaction, such as Na-phosphate, Na-citrate or Na-acetate as well as EDTA interfere with the MALDI sample preparation; presumably they disturb the matrix crystallization and/or analyte incorporation. Tris-HCl or ammonium salt buffers, in contrast, are MALDI compatible (Shaler, T. A. *et al.* (1996) Anal. Chem. 68:576-579). Moreover, alkaline salts in the sample lead to the formation of a heterogenous mixture of multiple salts of the analyte, a problem increasing with increasing number of phosphate groups. Such mixtures result in loss of mass resolution and accuracy as well as signal-to-noise ratio (Little, D. P. *et al.* (1995) Proc. Natl. Acad. Sci. U.S.A. 92:2318-2322; Nordhoff, E., Cramer, R. Karas, M., Hillenkamp, F., Kirpekar, F., Kristiansen, K. and Roepstorff, P. (1993) Nucleic Acids Res., 21, 3347-3357). Therefore, RNase digestions were carried out under somewhat modified conditions compared to the ones described in the literature. They are summarized above in table VII. For RNase T<sub>1</sub>, A, CL<sub>3</sub> and Cusativin, Tris-HCl (pH 6-7.5) was used as buffer. 20 mM DAC provides the pH of 5, recommended for maximum activity of RNases U<sub>2</sub> and PhyM. The concentration of 10-20 mM of these compounds were found to not interfere significantly with the MALDI analysis. To examine the specificity of the selected ribonucleases under these conditions, three synthetic 20-25mer RNA molecules with different nucleotide sequences were digested.

-177-

The MALDI-MS spectra of Figure 77 shows five different cleavage patterns (A-E) of a 25 nt RNA obtained after partial digestion with RNases T<sub>1</sub>, U<sub>2</sub>, PhyM, A, and alkaline hydrolysis. These spectra were taken from aliquots which were removed from the assay after empirically  
5 determined incubation times, chosen to get an optimum coverage of the sequence. As the resulting samples were not fractionated prior to mass spectrometric analysis, they contain all fragments generated at that time by the respective RNases. In practice, uniformity of the cleavages, can be affected by a preferential attack on the specific phosphodiester bonds  
10 (Donis-Keller, H., Maxam, A. M., and Gilbert, W. (1977) *Nucleic Acids Res.*, 4, 1957-1978; Donis-Keller, H. (1980) *Nucleic Acids Res.*, 8 3133-3142). The majority of the expected fragments are indeed observed in the spectra. It is also worth noting that for the reaction protocols as used, correct assignment of all fragment masses is only possible, if a 2',  
15 3'-cyclic phosphate group is assumed. It is well known that such cyclic phosphates are intermediates in the cleavage reaction and get hydrolyzed in a second, independent and slower reaction step involving the enzyme (Richards, F. M., and Wycoff, H. W. in *The Enzymes* Vol. 4, 3rd Ed., (ed. Boyer, P.D.) 746-806 (1971, Academic Press, New York); Heinemann, U  
20 and W. Saenger (1985) *Pure Appl. Chem.* 57, 417-422; Ikehara, M. et al., (1987) *Pure Appl. Chem.* 59-965-968) Vreslow, R. and Xu, R. (1993) *Proc. Nat. Acad. Sci. USA*, 90, 1201-1207). In a few cases different fragments have equal mass or differ by as little as 1 Dalton., In these cases, mass peaks cannot unambiguously be assigned to one or the  
25 other fragments. Digestion of two additional different 20 nt RNA samples was, therefore, performed (Hahner, S., Kirpekar, F., Nordoff, E., Kristiansen, K., Roepstorff, P. and Hillenkamp, F. (1996) *Proceedings of the 44th ASMS Conference on Mass Spectrometry*, Portland, Oregon) in order to sort out these ambiguities. For all samples tested, the selected

-178-

ribonucleases appear to cleave exclusively at the specified nucleotides leading to fragments arising from single as well as multiple cleavages.

In Figure 77, peaks, indicating fragments containing the original 5'-terminus, are marked by arrows. All non marked peaks can be assigned to internal sequences or those with retained 3'-terminus. For a complete sequence all possible fragments bearing exclusively either the 5'- or the 3'-terminus of the original RNA would suffice. In practice, the 5'-fragments are better suited for this purpose, because the spectra obtained after incubation of all three synthetic RNA samples contain the nearly complete set of originals of 5'-ions for all different RNases (Hahner, S., Kirpekar, F., Nordoff, E., Kristiansen, K., Roepstorff and Hillenkamp, F. (1996) *Proceedings of the 44th ASMS Conference on Mass Spectrometry*, Portland, Oregon). Internal fragments are somewhat less abundant and fragments containing the original 3'-terminus appear suppressed in the spectra. In agreement with observations reported in the literature (Gupta, R. C. and Randerath, K. (1977) *Nucleic Acids Res.*, 4, 1957-1978), cleavages close to the 3'-terminus were partially suppressed in partial digests of the RNA 25 mer by RNase T<sub>1</sub> and U<sub>2</sub> (even if they are internal or contain the original 5'-terminus). Fragments from such cleavages appear as weak and poorly resolved signals in the mass spectra.

For larger RNA molecules secondary structure is known to influence the uniformity of the enzymatic cleavages (Donis-Keller, H., Maxam, A. M. and Gilbert, A. (1977) *Nucleic Acids Res.* 8, 3133-3142). This can, in principle be, overcome by altered reaction conditions. In assay solutions containing 5-7 M urea, the activity of RNases such as T<sub>2</sub>, U<sub>2</sub>, A, Cl<sub>3</sub>, and PhyM is known to be retained (Donis-Keller, H., Maxam, A. M. and Gilbert, W. (1977) *Nucleic Acids Res.*, 4, 2527-2537; Boguski, M. S., Hieter, P.A., and Levy, C. C. (1980) *J. Biol. Chem.*, 255,



-179-

2160-2163; Donis-Keller, H. (1980) *Nucleic Acids Res.*, **8**, 3133-3142, while RNA is sufficiently denatured. UV-MALDI-analysis with 3-HPA as matrix is not possible under such high concentrations of urea in the sample. Up to a concentration of 2 M urea in the reaction buffer, MALDI  
5 analysis of the samples was still possible although significant changes in matrix crystallization were observed. Spectra of the RNA 20 mer (sample B), digested in the presence of 2 M urea still resembled those obtained under conditions listed in Table VII.

Digestion by RNases which exclusively recognize one nucleobase  
10 is desirable to reduce the complexity of the fragment patterns and thereby facilitate the mapping of the respective nucleobase. RNases CL<sub>3</sub> and cursavitin are enzymes reported to cleave at cytidylic acid residues. Upon limited RNase CL<sub>3</sub> and cursativin digestion of the RNA-20mer (sample B) under non-denaturing conditions, fragments corresponding to  
15 cleavages at cytidylic residues were indeed observed (Figure 78). Similar to the data reported so far (Boguski, M. S., Heiter, P. A. and Levy, C. C. (1980) *J. Biol. Chem.*, **255**, 2160-2163; Rojo, M. A., Arias, F. J., Iglesias, R., Ferreras, J. M., Munoz, R., Escarmis, C., Soriano, F., Llopez-Fando, Mendez, E., and Girbes, T. (1994) *Planta*, **194**, 328-338). The  
20 degradation pattern in Figure 78, however, reveals that not every cytidine residue is recognized, especially for neighboring C residues. RNase CL<sub>3</sub> is also reported to be susceptible to the influence of secondary structure (Boguski, M. S., Heiter, P. A., and Levy, C. C. (1980) *J Biol. Chem.*, **255**, 2160-2163), but for RNA of the size  
25 employed in this study, such an influence should be negligible.

Therefore, unrecognized cleavage sites in this case can be attributed to a lack of specificity of this enzyme. To confirm these data, a further RNase CL<sub>3</sub>-digestion was performed with the RNA 20mer (sample C). As a result of the sequence of this analyte, all three linkages containing

-180-

cytidylic acid were readily hydrolyzed, but additional cleavages at uridylic acid residues were detected as well. Since altered reaction conditions such as increased temperature (90°C), various enzyme to substrate ratios, and addition of 2M urea did not result in a digestion of the expected specificity, application of this enzyme to sequencing was not pursued further. Introduction of a new cytidine-specific ribonuclease, cusativin, isolated from dry seeds of *Cucumis sativus* L. looked promising for RNA sequencing (Rojo, M. A., Arias, F. J., Iglesias, R., Ferreras, J. M., Munoz, R., Escarmis, C., Soriano, F., Llopez-Fando, J., Mendez, E. and Girbes, T. (1994) *Planta*, 194, 328-338). As shown in Figure 78, not every cytidine residue was hydrolyzed and additional cleavages occurred at uridylic acid residues for the recommended concentration of the enzyme. RNases CL<sub>3</sub> and cusativin will, therefore not yield the desired sequence information for mapping of cytidine residues and their use was not further pursued. The distinction of pyrimidine residues can be achieved, however, by use of RNases with multiple specificities, such as *Physarum polycephalum* RNase (cleaves ApN, UpN) and pancreatic RNase A (cleaves UpN, CpN) (see Figure 77). All 5'-terminus fragments, generated by the monospecific RNase U<sub>2</sub> and apparent in the spectrum of Figure 77C were also evident in the spectrum of the RNase PhyM digest (Figure 77D). Five of the six uridylic cleavage sites could, this way, be uniquely identified by this indirect method. In a next step, the knowledge of the uridine cleavage sites was used to identify sites of cleavage of cytidylic acid residues in the spectrum recorded after incubation with RNase A (Figure 77E), again using exclusively ions containing the original 5'-terminus. Two of the four expected cleavage sites were identified this way. A few imitations are apparent from these spectra, if only the fragments containing the original 5'-terminus are used for the sequence determination. The first two nucleotides usually escape

-181-

the analysis, because their signals get lost in the low mass matrix background. Because of this, the corresponding fragments are missing in the spectra of the U- and C-specific cleavages. Large fragments with cleavage sites close to the 3'-terminus are often difficult to identify, particularly in digests with RNases T<sup>1</sup> and U<sub>2</sub>, because of their low yield (vide supra) and the often strong nearby signal of the non-digested transcript. Accordingly the cleavages in position 22 and 23 do not show up in the spectrum of the G-specific RNase T, (Figure 77A) and the cleavage site 24 cannot be identified from the spectra of the U<sub>2</sub> and PhyM digests (Figures 77 C and D). Also site 16 and 17 with two neighboring cytidilic acids cannot be identified in the RNase A spectrum of Figure 77E. These observations demonstrate that a determination of exclusively the 5'-terminus fragments may not always suffice and the information contained in the internal fragments may be needed for a full sequence analysis.

Finally, limited alkaline hydrolysis provides a continuum of fragments (Figure 77B), which can be used to complete the sequence data. Again, the spectrum is dominated by ions of fragments containing the 5'-terminus, although the hydrolysis should be equal for all phosphodiester bonds. As was true for the enzymatic digests, correct mass assignments requires one to assume that all fragments have a 2', 3'-cyclic phosphate. The distribution of peaks, therefore, resembles that obtained after a 3'-exonuclease digest (Pieles, U., Zurcher, W., Schar, M. and Moser, H. E., (1993) *Nucleic Acids Res.*, 21, 3191-3196; Nordhoff, E. et al. (1993) Book of Abstracts, 13<sup>th</sup> Internat. Mass Spectrom. Conf., Budapest p. 218; Kirpekar, F., Nordhoff, E., Kristiansen, K., Roepstorff, P., Lezius, A. Hahner, S., Karas, M. and Hillenkamp, F. (1994) *Nucleic Acid Res.*, 22, 3866-3870). In principle, the alkaline hydrolysis alone could, therefore, be used for a complete sequencing. This is, however,

only possible for quite small oligoribonucleotides, because larger fragment ions, differing in mass by only a few mass units will not be resolved in the spectra and the mass of larger ions cannot be determined with the necessary accuracy of better than 1 Da, even if peaks are partially or fully resolved. The interpretation of the spectra particularly from digests of unknown RNA samples is substantially simplified, if only the fragments containing the original 5'-terminus are separated out prior to the mass spectrometric analysis. A procedure for this approach is described in the following section.

- 10        *Separation of 5'-biotinylated fragments.* Streptavidin-coated magnetic beads (Dyna) were tested for the extraction of fragments containing the original 5'-terminus from the digests. Major features to be checked for this solid-phase approach are the selective immobilization and efficient elution of biotinylated species. In preliminary experiments, a 5'-biotinylated DNA (19 nt) and streptavidin were incubated and MALDI analyzed after standard preparation. Despite the high affinity of the streptavidin-biotin interaction, the intact complex was not found in the MALDI spectra. Instead, signals of the monomeric subunit of streptavidin and the biotinylated DNA were detected. Whether the complex dissociates in the acidic matrix solution (pK<sub>a</sub> 3) or during the MALDI desorption process, is not known. Surprisingly, if the streptavidin is immobilized on a solid surface such as magnetic beads, the same results are not observed. A mixture of two 5'-biotinylated DNA samples (19 nt and 27 nt) and two unlabeled DNA sequences (12 nt and 22 nt) were incubated with the beads. The beads were extracted and carefully washed before incubation in the 3-HPA MALDI matrix. No analyte signals could be obtained from these samples. To test whether the biotinylated species had been bound to the beads altogether, elution from the extracted and washed beads was performed by heating at 90°C in

-183-

the presence of 95% formamide. This procedure is expected to denature the streptavidin, thereby breaking the streptavidin/biotin complex. Figure 79B shows the expected signals of the two biotinylated species, proving that release of the bound molecules in the MALDI process is the problem rather than the binding of the beads; Figure 79A shows a spectrum of the same sample after standard preparation, showing signals of all four analytes as a reference. Complete removal of the formamide after the elution and prior to the mass spectrometric analysis was found to be important, otherwise crystallization of the matrix is disturbed. Mass resolution and the signal-to-noise ratio in spectrum 79B are comparable to those of the reference spectrum. These results testify to the specificity of the streptavidin-biotin interaction, since no or only minor signals of the non-biotinylated analyte were detected after incubation with the Dynal beads. Increased suppression of nonspecific binding was reported through an addition of the detergent Tween-20 to the binding buffer (Tong, X. and Smith, L. M. (1992) *Anal Chem.*, 64, 2672-2677). Although this effect could be confirmed in this study, peak broadening affected the quality of the spectra due to remaining amounts of the detergent. The necessity of an elution step as a prerequisite for detection of the captured biotinylated species can be attributed to a stabilizing effect of the complex by the immobilization of the streptavidin to the magnetic beads.

For practical application of this solid phase method to sequencing a maximum efficiency of binding and elution of biotinylated species is of prime importance. Among a variety of conditions investigated so far, addition of salts such as EDTA gave best results in the case of DNA sequencing by providing ionic strength to the buffer (Tong, X. and Smith, L. M. (1992) *Anal Chem.*, 64, 2672-2677). To examine such an effect on the solid-phase method, several salt additives were tested for the

-184-

binding and elution of the 5'-biotinylated RNA *in vitro* transcript (49 nt). The results are shown in Figure 80. Judging from the relative intensity, signal-to-noise ratio, and resolution of the respective signals, a 95% formamide solution containing 10 mM CDTA (Figure 80D) is most efficient for the binding/elution. Since CDTA acts as a chelating agent for divalent cation, formation of proper secondary and tertiary structure of the RNA is prevented. An improved sensitivity and spectral resolution has been demonstrated under such conditions for the analysis of RNA samples by electrospray mass spectrometry (Limbach, P. A., Crain, P. F. and McCloskey, J.A. (1995) *J Am. Soc. Mass. Spectrom.*, 6, 27-39). The improvement in the MALDI analysis is actually not very significant compared to the spectrum obtained for the solution containing formamide alone (Figure 81b), but the reproducibility for spectra of good quality was substantially improved for the CDTA/formamide solution. Thus in addition to the improved binding/elution, this additive may also improve the incorporation of the analyte into the matrix crystals. Unfortunately, a striking signal broadening on the high mass side was observed in case of formamide solutions containing EDTA, CDTA or 25% ammonium hydroxide. Since this effect is most prominent in case of 25% ammonium hydroxide and this agent was also used for adjusting EDTA and CDTA to their optimum pH, a pronounced  $\text{NH}_3$  adduct ion formation can be assumed.

The applicability of streptavidin-coated magnetic beads separation to RNA sequencing was demonstrated for the Rnase  $\text{U}_2$  digest of the 5'-biotinylated RNA *in vitro* transcript (49 nt) (Figure 81). The entire fragment pattern obtained after incubation with Rnase  $\text{U}_2$  is shown in spectrum 81A. Separation of the biotinylated fragments reduces the complexity of the spectrum (Figure 81B) since only 5'-terminal fragments are captured by the beads. The signals in the spectrum are broadened

-185-

and the increased number of signals in the low mass range indicate that even after stringent washing of the beads, some amounts of buffer and detergent used for the binding and elution remained. Further improvements of the method are, therefore, needed. Another possible  
5 strategy for application of the magnetic beads is the immobilization of the target RNA prior to RNase digestion by an elution of the remaining fragments for further analysis. Cleavage of the RNA was impeded in this case, as evidenced by a prolonged reaction time for the digest under otherwise identical reaction conditions.

10

**EXAMPLE 22****Parallel DNA Sequencing Mutation Analysis and Microsatellite Analysis Using Primers with Tags and Mass Spectrometric Detection**

This EXAMPLE describes specific capturing of DNA products generated in DNA analysis. The capturing is mediated by a specific tag  
15 (5 to 8 nucleotides long) at the 5' end of the analysis product that binds to a complementary sequence. The capture sequence can be provided by a partially double stranded oligonucleotide bound to a solid support. Different DNA analysis (e.g., sequencing, mutation, diagnostic, microsatellite analysis) can be carried out in parallel, using, for example,  
20 a conventional tube or microtiter plate (MTP). The products are then specifically captured and sorted out via the complementary identification sequence on the tag oligonucleotide. The capture oligonucleotide can be bound onto a solid support (e.g., silicon chip) by a chemical or biological bond. Identification of the sample is provided by the predefined position  
25 of the capture oligonucleotide. Purification, conditioning and analysis by mass spectrometry are done on solid support. This method was applied for capturing specific primers that had a 6 base tag sequence.

**MATERIALS AND METHODS***Genomic DNA.*

-186-

Genomic DNA was obtained from healthy individuals.

*PCR Amplification*

PCR amplifications of part of the  $\beta$ -globin gene were established using  $\beta$ 2 d(CATTGCTTCTGACACAACT Seq. ID. No. 66) as forward  
5 primer and  $\beta$ 11 d(TCTCTGTCTCCACATGCCAG Seq. ID. No. 67) as reverse primer. The total PCR volume was 50  $\mu$ l including 200 ng genomic DNA, 1U Taq-polymerase (Boehringer-Mannheim, Cat# 159594), 1.5 mM  $MgCl_2$ , 0.2 mM dNTPs (Boehringer-Mannheim, Cat# 1277049), and 10 pmol of each primer. A specific fragment of the  $\beta$ -  
10 globin gene was amplified using the following cycling conditions: 5 min @ 94°C followed by 40 cycles of 30 sec @ 94°C, 45 sec @ 53°C, 30 sec @ 72°C, and a final extension of 2 min @ 72°C. Purification of the amplified product and removal of unincorporated nucleotides was carried out using the QIAquick purification kit (Qiagen, Cat 28104). One fifth of  
15 the purified product was used for the primer oligo base extension (PROBE) or sequencing reactions, respectively.

*Primer oligo base extension (PROBE) and sequencing reactions*

Detection of putative mutations in the human  $\beta$ -globin gene at codon 5 and 6 and at codon 30 and in the IVS-1 donor site, respectively,  
20 was done in parallel (FIGURE 82A).  $\beta$ -TAG1 (GTCGTCCCATGGTGCACCTGACTC Seq. ID. No. 68) served as primer to analyze codon 5 and 6 and  $\beta$ -TAG2 (CGCTGTGGTGAGGCCCTGGGCA Seq. ID. No. 69) for the analyses of codon 30 and the IVS-1 donor site. The primer oligo base extension (PROBE) reaction was done by cycling,  
25 using the following conditions: final reaction volume was 20  $\mu$ l,  $\beta$ -TAG1 primer (5 pmol),  $\beta$ -TAG2 primer (5 pmol), dCTP, dGTP, dTTP, (final concentration each 25  $\mu$ M), ddATP (final concentration 100  $\mu$ M) dNTPs and ddNTPs purchased from Boehringer-Mannheim, Cat# 1277049 and 1008382), 2  $\mu$ l of 10x ThermoSequence buffer and 2.5 U



-187-

ThermoSequenase (Amersham, CAT#E79000Y). The cycling program was as follows: 5 min @ 94°C, 30 sec @ 53°C, 30 sec @ 72°C and a final extension step for 8 min @ 72°C. Sequencing was performed under the same conditions except that the reaction volume was 25  $\mu$ l and the concentration of nucleotides was 250  $\mu$ M for ddNTP.

*Capturing using TAG sequence and sample preparation*

The capture oligonucleotides cap-tag1 d(GACGACGACTGCTACCTGACTCCA Seq ID No. 70) and cap-tag2 d(ACAGCGGACTGCTACCTGACTCCA Seq ID No. 71), respectively, were annealed to equimolar amounts of uni-as d(TGGAGTCAGGTAGCAGTC Seq ID No. 72) (FIGURE 82A). Each oligonucleotide had a concentration of 10 pmol/ $\mu$ l in ddH<sub>2</sub>O and incubated for 2 min @ 80°C and 5 min @ 37°C. This solution was stored at -20°C and aliquots were taken. 10 pmol annealed capture oligonucleotides were bound to 10  $\mu$ l paramagnetic beads coated with streptavidin (10 mg/ml; Dynal, Dynabeads M-280 streptavidin Cat# 112.06) by incubation for 30 min @ 37°C. Beads were captured and the PROBE or sequencing reaction, respectively, was added to the capture oligonucleotides. To facilitate binding of  $\beta$ -TAG1 and  $\beta$ -TAG2, respectively, the reaction was incubated for 5 min @ 25°C and for 30 min @ 16°C. The beads were washed twice with ice cold 0.7 M NH<sub>4</sub> Citrate to wash away unspecific bound extension products and primers. The bound products were dissolved by adding 1  $\mu$ l DDH<sub>2</sub>O and incubation for 2 min @ 65°C and cooling on ice. 0.3  $\mu$ l of the sample were mixed with 0.3  $\mu$ l matrix solution (saturated 3-hydroxypicolinic acid, 10% molar ratio ammonium-citrate in acetonitrile/water (50/50. v/v)) and allowed to air dry. The sample target was automatically introduced into the source region of an unmodified Perspective Voyager MALDI-TOF operated in delay d extraction linear mode with 5 and 20 kV on the target and conversion dynod ,

-188-

respectively. Theoretical average molecular mass ( $M_r(\text{calc})$ ) were calculated from atomic compositions; reported experimental  $M_r(M_r(\text{exp}))$  values are those of the singly-protonated form.

## RESULTS

5        Specific capturing of a mixture of extension products by a short complementary sequence has been applied to isolate sequencing and primer oligo base extension (PROBE) products. This method was used for the detection of putative mutations in the human  $\beta$ -globin gene at codon 5 and 6 and at codon 30 and IVS-1 donor site, respectively  
10 (FIGURE 82A). Genomic DNA has been amplified using the primers  $\beta$ 2 and  $\beta$ 11. The amplification product was purified and the nucleotides separated. One fifth of the purified product was used for analyses by primer oligo base extension. To analyze both sites in a single reaction, primers,  $\beta$ -TAG1 and  $\beta$ -TAG2, were used respectively.  $\beta$ -TAG1 binds  
15 upstream of codons 5 and 6 and  $\beta$ -TAG2 upstream of codon 30 and the IVS-1 donor site. Extension of these primers was performed by cycling in the presence of ddATP and dCTP, dGTP and dTTP, leading to specific products, depending on the phenotype of the individual. The reactions were then mixed with the capture oligonucleotides. Capture  
20 oligonucleotides include the biotinylated capture primer cap-tag1 and cap-tag2, respectively. They have 6 bases at the 5' end, that are complementary to the 5' end of  $\beta$ -TAG1 and  $\beta$ -TAG2, respectively. Therefore, they specifically capture these primers and the extended products. By annealing a universal oligonucleotide (uni-as) to the capture  
25 oligonucleotide, the capture primer is transformed into a partially double stranded molecule where only the capture sequence stays single stranded (Figure 82). This molecule is then bound to streptavidin coated paramagnetic particles, to which the PROBE or sequencing reaction, respectively is added. The mixture was washed to bind only the

-189-

specifically annealed oligonucleotides. Captured oligonucleotides are dissolved and analyzed by mass spectrometry.

PROBE products of one individual (Fig. 83) show a small peak with a molecular mass of 7282.8 Da. This corresponds to the unextended  $\beta$ -TAG1 that has a calculated mass of 7287.8 Da. The peak at 8498.6 Da corresponds to a product, that has been extended by 4 bases. This corresponds to the wildtype situation. The calculated mass of this product is 8500.6 Da. There is no significant peak indicating a heterozygote situation. Furthermore only  $\beta$ -TAG1 and not  $\beta$ -TAG2 has been captured, indicating a high specificity of this method.

Analyses of what was bound to cap-tag2 (Figure 84) shows only one predominant peak with a molecular mass of 9331.5 Da. This corresponds to an extension of 8 nucleotides. It indicates a homozygous wildtype situation where the calculated mass of the expected product is 9355 Da. There is no significant amount of unextended primer and only  $\beta$ -TAG2 has been captured.

To prove that this approach is also suitable for capturing specific sequencing products, the same two primers  $\beta$ -TAG1 and  $\beta$ -TAG2, respectively, were used. The primers were mixed, used in one sequencing reaction and then sorted by applying the above explained method. Two different termination reactions using ddATP and ddCTP were performed with these primers (Figures 85 and 86, respectively). All observed peaks in the spectrograms correspond to the calculated masses in a wildtype situation.

As shown above, parallel analysis of different mutations (e.g., different PROBE primers) is now possible. Further, the described method is suitable for capturing specific sequencing products. Capturing can be used for separation of different sequencing primers out of one reaction tube/well, isolation of specific multiplex-amplified products, PROBE

-190-

products, etc. Conventional methods, like cycle sequencing, and conventional volumes can be used. A universal chip design permits the use of many different applications. Further, this method can be automated for high throughput.

5

**EXAMPLE 23****Deletion Detection by Mass-Spectrometry**

Various formats can be employed for mass spectrometer detection of a deletion within a gene. For example, molecular mass of a double standard amplified product can be determined, or either or both of the  
10 strands of a double stranded product can be isolated and the mass measured as described in previous examples.

Alternatively, as described herein, a specific enzymatic reaction can be performed and the mass of the corresponding product can be determined by mass spectrometry. The deletion size can be up to  
15 several tenths of bases in length, still allowing the simultaneous detection of the wildtype and mutated allele. By simultaneous detection of the specific products, it is possible to identify in a single reaction whether the individual is homozygous or heterozygous for a specific allele or mutation.

**20 MATERIALS AND METHODS***Genomic DNA*

Leukocyte genomic DNA was obtained from unrelated healthy individuals.

*PCR amplification*

25

PCR amplification of the target DNA was established and optimized to use the reaction products without a further purification step for capturing with streptavidin coated beads. The primers for target amplification and for PROBE reactions were as follows:

-191-

CKRA-F:d(CAG CTC TCA TTT TCC ATA C SEQ ID. NO. 73) and CKRA-R bio: d(AGC CCC AAG ATG ACT ATC SEQ ID. NO. 74). CKR-5 was amplified by the following program: 2 min @ 94°C, 45 seconds @ 52°C, 5 seconds @ 72°C, and a final extension of 5 minutes at 72°C.

5. The final volume was 50  $\mu$ l including 200 ng genomic DNA 1U Taq-polymerase (Boehringer-Mannheim, Cat # 1596594), 1.5 Mm  $MgCl_2$ , 0.2 Mm DNTPS (Boehringer-Mannheim, Cat # 1277049), 10 pmol of unmodified forward primers, and 8 pmol 5' biotinylated reverse primer.

*Capturing and Denaturation of Biotinylated Templates*

- 10 10  $\mu$ l paramagnetic beads coated with streptavidin (10 mg/ml; Dynal, Dynabeads M-280 streptavidin Cat # 112.06) in 5x binding solution (5M  $NH_4Cl$ , 0.3 M  $NH_4OH$ ) were added to 45  $\mu$ l PCR reaction (5  $\mu$ l of PCR reaction were saved for electrophoresis). After binding by incubation for 30 min. at 37°C the supernatant was discarded. Captured
- 15 templates were denatured with 50  $\mu$ l of 100 Mm NaOH for 5 min. at ambient temperature, washed once with 50  $\mu$ l 50 Mm  $NH_4OH$  and three times with 100  $\mu$ l 10 Mm Tris/Cl, Ph 8.0. The single stranded DNA served as templates for PROBE reactions.

*Primer Oligo Base Extension (PROBE) Reaction*

- 20 The PROBE reaction was performed using Sequence 2.0 (USB Cat # E70775Z including buffer). dATP/DGTP and ddTTP were supplied by Boehringer-Mannheim (Cat # 1277049 and 1008382). d(CAG CTC TCA TTT TCC ATA C (SEQ ID. NO. 73) was used as PROBE primer (Figure 87). The following solutions were added tot he beads: 3.0  $\mu$ l  $H_2O$ , 1.0  $\mu$ l
- 25 reaction buffer, 1.0  $\mu$ l PROBE primer (10 pmol) and incubated at 65°C for 5 minutes followed by 37°C for 10 min. Then 0.5  $\mu$ l DTT, 3.5  $\mu$ l DNTPS/ddntp each 50  $\mu$ M and 0.5  $\mu$ l Sequenase (0.8 U) wer added and incubated at 37°C for 10 min.

*T4 Treatment of DNA*

-192-

To generate blunt ended DNA, amplification products were treated with T4 DNA polymerase (Boehringer-Mannheim Cat# 1004786). The reactions were carried out according to the manufacturer's protocol for 20 min. at 11°C.

5 *Direct Size Determination of Extended Products*

To determine the size of the amplified product, MALDI-TOF was applied to one strand of the amplification product. samples were bound to beads, as described above, conditioned and denatured, as described below.

10 *DNA Conditioning*

After the PROBE reaction the supernatant was discarded and the beads were washed first in 50  $\mu$ l 700 mM  $\text{NH}_4$ -citrate and second 50  $\mu$ l 50 mM  $\text{NH}_4$ -citrate. The generated diagnostic products were removed for the template by heating the beads in 2  $\mu$ l  $\text{H}_2\text{O}$  at 80°C for 2 min.

15 The supernatant was used for MALDI-TOF analysis.

*Sample Preparation and Analysis with MALDI-TOF Mass Spectrometry*

Sample preparation was performed by mixing 0.6  $\mu$ l of matrix solution (0.7 M 3-hydroxypicolinic acid, 0.07 M dibasic citrate in 1:1

20  $\text{H}_2\text{O}:\text{CH}_3\text{CN}$ ) with 0.3  $\mu$ l of diagnostic PROBE products in water on a sample target and allowed to air dry. Up to 100 samples were spotted on a probe target disk for introduction into the source region of an unmodified Perspective Voyager MALDI-TOF instrument operated in linear mode with delayed extraction and 5 and 30 kV on the target and

25 conversion dynode, respectively. Theoretical average molecular mass ( $M_r(\text{calc})$ ) of analytes were calculated from atomic compositions, reported experimental  $M_r(M_r(\text{exp}))$  values are those of the singly-protonated form, determined using internal calibration with unextended primers in the case of PROBE reactions.

-193-

### *Conventional Analyses*

Conventional analyses were performed by native polyacrylamide gel electrophoresis according to standard protocols. The diagnostic products were denatured with formamide prior to loading onto the gels and stained with ethidium bromide or silver, respectively.

### **RESULTS**

The CKR-5 status of 10 randomly chosen DNA samples of healthy individuals were analyzed. Leukocyte DNA was amplified by PCR and an aliquot of the amplified product was analyzed by standard polyacrylamide gel electrophoresis and silver staining of the DNA (Figure 88). Four samples showed two bands presumably indicating heterozygosity for CKR-5, whereas the other 6 samples showed one band, corresponding to a homozygous gene (Figure 88). In the case where two bands were observed, they correspond to the expected size of 75 bp for the wildtype gene and 43 bp for the allele with the deletion (Figure 87). Where one band was observed, the size was about 75 bp which indicated a homozygous wildtype CKR-5 allele. One DNA sample derived from a presumably heterozygous one from a homozygous individual were used for all further analysis. To determine the molecular mass of the amplified product, DNA was subjected to matrix assisted laser desorption/ionization coupled with time of flight analysis (MALDI-TOF). Double stranded DNA, bound to streptavidin coated paramagnetic particles, was denatured and the strand released into the supernatant was analyzed. Figure 89A shows a spectrograph of a DNA sample, that was supposed to be heterozygous according to the result derived by polyacrylamide gel electrophoresis (Figure 88). The calculated mass of the sense strand for a wildtype gene is 23036 Da and for the sense strand carrying the deletion allele 13143 (Figure 87 and Table VI). Since many thermostable polymerases unspecifically add an adenosine to the

-194-

3' end of the product, those masses were also calculated. They are 23349 and 13456 Da. The masses of the observed peaks (Figure 89A) are 23119 Da, which corresponds to the calculated mass of a wildtype DNA strand where an adenosine has been added (23349 Da). Since no  
5 peak with a mass of about 23036 Da was observed, the polymerase must have qualitatively added adenosine. Two peaks, which are close to each other, have a mass of 13451 and 13137 Da. This corresponds to the calculated masses of the allele, with the 32bp deletion. The higher mass peak corresponds to the product, where adenosine has been added  
10 and the lower mass peak to the one without the unspecific adenosine. Both peaks have about the same height, indicating that to about half of the product adenosine has been added. The peak with a mass of 11682 Da is a doubly charged molecule of the DNA corresponding to 23319 Da ( $2 \times 11682 \text{ Da} = 23364 \text{ Da}$ ). The peaks with masses of 6732 and  
15 6575 Da are doubly charged molecules of the one with masses of 13451 and 13137 Da and the peak with 7794 Da corresponds to the triply charged molecule of 23319Da. Multiple charged molecules are routinely identified by calculation. Amplified DNA derived from a homozygous individual shows in the spectrograph (Figure 89C) one peak with a mass  
20 23349.6 and a much smaller peak with a mass of 23039.9 Da. The higher mass peak corresponds to DNA resulting from a wildtype allele with an added adenosine, that has a calculated mass of 23349 Da. The lower mass peak corresponds to the same product without adenosine. Three further peaks with a mass of 11686, 7804.6 and 5852.5 Da  
25 correspond to doubly, triply and quadruply charged molecules.

The unspecific added adenine can be removed from the amplified DNA by treatment of the DNA and T4 DNA polymerase. DNA derived from a heterozygous and a homozygous individual was analyzed after T4 DNA polymerase treatment. Figure 89B shows the spectrograph



-195-

derived from heterozygous DNA. The peak corresponding to the wildtype strand has a mass of 23008 Da indicating that the added adenine had been removed completely. The same is observed for the strand with a mass of 13140 Da.

- 5 The other three peaks are multiply charged molecules of the parent peaks. The mass spectrograph for the homozygous DNA shows one peak that has a mass of 23004 Da, corresponding to the wildtype DNA strand without an extra adenine added. All other peaks are derived from multiply charged molecules of this DNA. The amplified products can be
- 10 analyzed by direct determination of their masses, as described above, or by measuring the masses of products, that are derived from the amplified product in a further reaction. In this "primer oligo base extension (PROBE)" reaction, a primer that can be internal, as it is in the nested PCR, or identical to one of the PCR primers, is extended for just a few
- 15 bases before the termination nucleotide is incorporated. Depending on the extension length, the genotype can be specified. CKRA-F was used as a PROBE primer, and dATP/dGTP and ddTTP as nucleotides. The primer extension is AGT in case of a wildtype template and AT in case of the deletion (Figure 87). The corresponding masses are 6604 Da for the
- 20 wildtype and 6275 Da for the deletion, respectively. PROBE was applied to two standard DNAs. The spectrograph (Figure 90A) shows peaks with masses of 6604 Da corresponding to the wildtype DNA and at 6275 Da corresponding to the CKR-5 deletion allele (Table VIII). The peak at a mass of 5673 Da corresponds to CKRA-F (calculated mass of 5674 Da).
- 25 Further samples were analyzed in analogous way (Figure 90B). It is unambiguously identified as homozygous DNA, since the peak with a mass of 6607 Da corresponds to the wildtype allele and the peak with a mass of 5677 Da to the unextended primer. No further peaks were observed.

-196-

The example demonstrates that deletion analysis can be performed by mass spectrometry. As shown herein, the deletion can be analyzed by direct detection of single stranded amplified products, or by analysis of specifically generated diagnostic products (PROBE). In addition, as shown in the following Example 26, double stranded DNA amplified products can be analyzed.

Size	Calculated Mass	Measured Mass
wildtype w/o A	23036	23039/23009/23004
wildtype with A	23349	23319/23350
deletion w/o A	13143	13137/13139
deletion with A	13456	13451
PROBE		
wildtype	6604	6604/6608
deletion	6275	6275

All masses are in Dalton.

#### EXAMPLE 24

##### Pentaplex tc-PROBE

##### SUMMARY

The multiplexing of thermocycling primer oligo base extension (tc-PROBE) was performed using five polymorphic sites in three different apolipoprotein genes, which are thought to be involved in the pathogenesis of atherosclerosis. The apolipoprotein A IV gene (codons 347 and 360), the apolipoprotein E gene (codons 112 and 158), and the apolipoprotein B gene (codon 3500) were examined. All mass spectra were easy to interpret with respect to the five polymorphic sites.

-197-

**MATERIALS AND METHODS****PCR Amplification**

Human leukocytic genomic DNA was used for PCR. Listed below are the primers used for the separated amplification of portions of the

**5 Apo A IV, Apo E and the Apo B genes:**

5	Apo A IV:	A347F:	5'-CGA GGA GCT CAA GGC CAG AAT-3' (SEQ ID NO. 75)
		A360 R-2-bio:	*5'-CAG GGG CAG CTC AGC TCT C-3' (SEQ ID NO. 76)
	Apo E:	ApoE-F:	5'-GGC ACG GCT GTC CAA GGA-3' (SEQ ID NO. 77)
		ApoE-R bio:	*5'-AGG CCG CGC TCG GCG CCC TC-3' (SEQ ID NO. 78)
10	Apo B:	ApoB-F2 bio:	*5'-CTT ACT TGA ATT CCA AGA GC-3' (SEQ ID NO. 79)
		Apo B-R:	5'-GGG CTG ACT TGC ATG GAC CGG A-3' (SEQ ID NO. 80)

\* *biotinylated*

Taq polymerase and 10x buffer were purchased from Boehringer-Mannheim (Germany) and dNTPs for Pharmacia (Freiburg, Germany). The total PCR reaction volume was 50  $\mu$ l including 10 pmol of each primer and 10% DMSO (dimethylsulfoxide, Sigma) (no DMSO for the PCR of the Apo B gene), with ~200 mg of genomic DNA used as template and a final dNTP concentration of 200  $\mu$ M. Solutions were heated to 80°C before the addition of 1U Taq polymerase; PCR conditions were: 5 min at 95°C, followed by 2 cycles 30 sec 94°C, 30 sec 62°C, 30 sec 72°C, 2 cycles 30 sec 94°C 30 sec 58°C, 30 sec 72°C, 35 cycles of 30 sec at 94°C, 30 sec at 56°C, 30 sec at 72°C, and a final extension time of 2 min at 72°C. To remove unincorporated primers and nucleotides, amplified products were purified using the "QIAquick" (Qiagen, Germany) kit, with elution of the purified products in 50  $\mu$ L of TE buffer (10 mM Tris-HCl, 1mM EDTA, pH 8.0).

-198-

*Binding of the amplified product on beads*

10  $\mu$ l of each purified amplified product was bound to 5  $\mu$ l DynaBeads (Dyna, M-280 Streptavidin) and denatured according to the protocol from Dynal. For the pentaplex tc-PROBE reaction the three  
5 different amplified product (bound on the beads) were pooled.

**Tc-PROBE**

For the PROBE reaction the following primers were used:

- (Apo A) P347: 5'-AGC CAG GAC AAG-3' (SEQ ID NO. 81)  
(Apo A) P360: 5'-ACA GCA GGA ACA GCA-3' (SEQ ID NO. 82)  
10 (Apo E) P112: 5'-GCG GAC ATG GAG GAC GTG-3' (SEQ ID NO. 83)  
(Apo E) P158: 5'-GAT GCC GAT GAC CTG CAG AAG-3' (SEQ ID NO. 84)  
(Apo B) P3500: 5'-GTG CCC TGC AGC TTC ACT GAA GAC-3' (SEQ ID NO. 85)

- 15 The tc-PROBE was carried out in a final volume of 25  $\mu$ l containing 10 pmol of each primer listed above, 2.5 U Thermoquenase (Amersham), 2.5  $\mu$ L Thermoquenase buffer, and 50  $\mu$ M dTTP (final concentrations) and 200  $\mu$ M of ddA/C/GTP, respectively. Tubes containing the mixture were placed in a thermocycler and subjected to the following cycling  
20 conditions: denaturation (94°C) the supernatant was carefully removed from the beads and 'desalted' by ethanol precipitation to exchange nonvolatile cations such as Na<sup>+</sup> and K<sup>+</sup> with NH<sub>4</sub><sup>+</sup>, which evaporated during the ionization process; 5  $\mu$ L 3M ammonium acetate (pH 6.5) 0.5  $\mu$ L glycogen (10 mg/mL, Sigma), 25  $\mu$ L H<sub>2</sub>O, and 110  $\mu$ L absolute ethanol  
25 were added to 25  $\mu$ L PROBE supernatant and incubated for 1 hour at 4°C. After a 10 min. centrifugation at 13,000 X g, the pellet was washed in 70% ethanol and resuspended in 1  $\mu$ L 18 Mohm/cm H<sub>2</sub>O. A 0.35  $\mu$ L aliquot of resuspended DNA was mixed with 0.35  $\mu$ L matrix solution (0.7 M 3-hydroxypicolinic acid (3-HPA), 0.07 M ammonium

-199-

citrate in 1:1 H<sub>2</sub>O:CH<sub>3</sub>CN) on a stainless steel sample target disk and allowed to air dry preceding spectrum acquisition using the Thermo Bioanalysis Version 2000 MALDI-TOF operated in reflectron mode with 5 and 20 kV on the target and conversion dynode, respectively, Theoretical average molecular masses (M<sub>1</sub>(calc)) of the fragments were calculated from atomic compositions. External calibration generated from synthetic (ATCG)<sub>n</sub> oligonucleotide (3.6-18kDa) was used. Positive ion spectra from 1-37500 Da were collected.

## RESULTS

- 10 Table VIII shows the calculated molecular masses of all possible extension products including the mass of the primer itself. Fig. 91 shows a respective MALDI-TOP MS spectra of a tc-PROBE using three different templates and 5 different PROBE primers simultaneously in ne reaction. Comparison of the observed and calculated masses (see table 15 VIII) allows a fast genetic profiling of various polymorphic sites in an individual DNA sample. The sample presented in Figure 91 is homozygous for threonine and glutamine at position 347 and 360, respectively, in the apolipoprotein A IV gene, bears the epsilon 3 allele homozygous in the apolipoprotein E gene, and is also homozygous at the 20 codon 3500 for arginine in the apolipoprotein B gene.

TABLE VIII

	SEQ ID	mass	allele
<b>Apolipoprotein A IV</b>			
5'-AGCCAGGACAAG-3' (347)	86	3688.40	unextended primer
25 5'-AGCCAGGACAAGTC-3'	87	4265.80	347Ser
5'-AGCCAGGACAAGA-3'	88	3985.60	347Thr
5'-ACAGCACCAACAGCA-3'(360)	89	4604.00	unextended primer
5'-ACAGCAGGAACAGCATC-3'	90	5181.40	360His

-200-

	SEQ ID	mass	allele
5'-ACAGCAGGAACAGCAG-3' (112)	91	4917.20	360Gln
<b>Apolipoprotein E</b>			
5'-GCGGACATGGAGGACGTG-3' (112)	92	5629.60	unextended primer
5'-GCGGACATGGAGGACGTGGC-3'	93	6247.00	112Cys
5'-GCGGACATGGAGGACGTGC-3'	94	5902.80	112Arg
5'-GATGCCGATGACCTGCAGAAG-3'(158)	95	6480.20	unextended primer
5'-GATGCCGATGACCTGCAGAAGC-3'	96	6753.40	158Arg
5'-GATGCCGATGACCTGCAGAAGTG-3'	97	7097.60	158Cys
<b>Apolipoprotein B-100</b>			
5'-GTGCCCTGCAGCTTCACTGAAGAC-3' (3500)	98	7313.80	unextended primer
5'-GTGCCCTGCAGCTTCACTGAAGACTG-3'	99	7931.20	3500Gln
5'-GTGCCCTGCAGCTTCACTGAAGACC-3'	100	7587.00	3500Arg

**EXAMPLE 25**

**Sequencing Exons 5 to 8 of the p53 Gene by MALDI-TOF Mass Spectrometry**

**MATERIALS & METHODS**

- Thirty-five cycles of PCR reactions were performed in a 96 well microliter plate with each well containing a total volume of 50  $\mu$ l including 200 ng genomic DNA, 1 unit Taq DNA polymerase, 1.5 mM Mg C<sub>1</sub>, 0.2mM dNTPx, 10 pmol of the forward primer and 6 or 8 of the biotinylated reverse primer. The sequences of PCR primers prepared according to established chemistry (N.D. Sinha, J. Biernat, H. Kter, Tetrahed. Lett. 24:5843-5846 (1983) are as follows: exon 5:d(biotin-TATCTGTTCACTTGTGCC SEQ ID NO. 101) and d(biotin-CAGAGGCCTGGGGACCCTG SEQ ID NO.102); exon 6:

-201-

D(ACGACAGGGCTGGTTGCC SEQ ID NO. 103) and d(biotin-  
ACTGACAACCACCCTTAAC SEQ ID NO. 104); exon 7:

d(CTGCTTGCCACAGGTCTC SEQ ID NO. 105) and d(biotin-  
CACAGCAGGCCAGTGTGC SEQ ID NO. 106; exon 8:

5. d(GGACCTGATTTTCCTTACTG SEQ ID NO. 107) and d(biotin-  
TGAATCTGAGGCATAACTG SEQ ID NO. 108).

To each well of the 96-well microliter plate containing unpurified amplified product, 0.1 mg of paramagnetic streptavidin beads (Dynal) in 10  $\mu$ l of 5 x binding solution (5 M  $\text{NH}_4\text{OH}$ ) was added and incubated at 3  
10 7°C for 30 min.

Then beads were treated with 0.1 M NaOH at room temperature for 5 min followed by one wash with 50 mM  $\text{NH}_4\text{OH}$  at room temperature for 5 min followed by one wash with 50 mM Tris-HCl.

Four dideoxy termination reactions were carried out in separate  
15 wells of the microliter plate. A total of 84 reactions (21 primers x 4 reactions/primer) can be performed in a single microliter plate. To each well containing immobilized single-stranded template, a total volume of 10  $\mu$ l reaction mixture was added including 1x reaction buffer, 10 pmol of sequencing primer, 250 mM of dNTPs, 25 mM of one of the ddNTPs,  
20 and 1-2 units of Thermosequenase (Amersham). Sequencing reactions were carried out on a thermal cycler using non-cycling conditions: 80°C, 1 min, 50°C, 1 min, 50°C to 72°C, ramping 0.1 °/sec, and 72°C, 5 min. The beads were then washed with 0.7 M ammonium citrate followed by 0.05 M ammonium citrate. Sequencing products were then removed  
25 from beads by heating the beads to 80°C in 2  $\mu$ l of 50mM  $\text{NH}_4\text{OH}$  for 2 min. The supernatant was used for MALDI-TOF MS analysis.

Matrix was prepared as described in Kter, et al (Kter, H. *et al.*, Nature Biotechnol. 14: 1123-1128 (1996)). This saturated matrix solution was then diluted 1.52 times with pure water before use. 0.3  $\mu$ l

-202-

of the diluted matrix solution was then diluted 1.52 times with pure water before use. 0.3  $\mu$ l of the diluted matrix solution was loaded onto the sample target and allowed to crystallize followed by addition of 0.3  $\mu$ l of the aqueous analyte. A Perseptive Voyager DE mass spectrometer was used for the experiments, and the samples were typically analyzed in the manual mode. The target and middle plate were kept at +18.2 kV for 200 nanoseconds after each laser shot and then the target voltage was raised to +20kV. the ion guide wire in the flight tube was kept at -2V. Normally, 250 laser shots were accumulated for each sample. The original spectrum was acquired under 500 MHz digitizing rate, and the final spectrum was smoothed by a 455 point average (Savitsky and Golay, (1964) Analytical Chemistry, 36:1627). Default calibration of the mass spectrometer was used to identify each peak and assign sequences. The theoretical mass values of two sequencing peaks were used to recalibrate each spectrum. (D.P. Little, T.J. Cornish, M.J. O'Donnell, A. Braun, R.J. Cotter, H. Kter, Anal. Chem., submitted).

## RESULTS

Alterations of the p53 gene are considered to be a critical step in the development of many human cancers (Greenblatt, et al., (1994) Cancer Res. 54, 4855--4878; C.C. Harris, (1996) J. Cancer, 73, 261-269; and D. Sidransky and M. Hollstein, (1996) Annu.Res.Med., 47,285-301). Mutations may serve as molecular indicators of clonality or as early markers of relapse in a patient with a previously identified mutation in a primary tumor (Hainaut, et al., (1997) Nucleic Acid Res., 25, 151-157). The prognosis of the cancer may differ according to the nature of the p53 mutations present (H.S. Goh et al., (1995) Cancer Res, 55, 5217-5221). Since the discovery of the p53 gene, more than 6000 different mutations have been detected. Exons 5-8 were selected as



-203-

sequencing targets where most of the mutations cluster (Hainaut et al. (1997) *Nucleic Acids Res.*, 25, 151-7).

Figure 96 schematically depicts the single tube process for target amplification and sequencing, which was performed, as described in detail in the Materials and Methods. Each of exon 5-8 of the p53 gene was PCR amplified using flanking primers in the intron region; the downstream primer was biotinylated. Amplifications of different exons were optimized to use the same cycling profile, and the products were used without further purification. PCR reactions were performed in a 96 well microliter plate and the product generated in one well was used as the template for one sequencing reaction. Streptavidin-coated magnetic beads were added to the same microliter plate and amplified products were immobilized. The beads were then treated with NaOH to generate immobilized single-stranded DNA as sequencing template. The beads were washed extensively with Tris buffer since remaining base would reduce the activity of sequencing enzyme.

A total of 21 primers were selected to sequence exon 5-8 of the p53 gene by primer walking. The 3'-end nucleotide of all the primers is located at the site where no known mutation exists. Four termination reactions were performed separately which resulted in a total of 84 sequencing reactions on the same PCR microliter plate. Non-cycling conditions were adopted for sequencing since streptavidin coated beads do not tolerate the repeated application of high temperature. Sequencing reactions were designed so that mt terminated fragments were under 70 nucleotides, a size range easily accessible by MALDI-TOF MS and yet long enough to sequence through the next primer binding site.

Thermequenase was the enzyme of choice since it could reproducibly generate a high yield of sequencing products in the desired mass range. After the sequencing reactions, the beads were washed with ammonium

-204-

ion buffers to replace all other cations. The sequencing ladders were then removed from the beads by heating in ammonium hydroxide solution or simply in water.

A sub-microliter aliquot of each of the 84 sequencing reactions  
5 was loaded onto one MS sample holder containing preloaded matrix. Figure 94 gives an example of sequencing data generated from one primer; four spectra are superimposed.

All sequencing peaks were well resolved in the mass range needed to read through the next sequencing primer site. Sometimes doubly  
10 charged peaks were observed which could be easily identified by correlating the mass to that of the singly charged ion. False stops generated by early termination of the enzymatic extension can be observed cle to the primer site. Since the mass resolution is high enough, it is easy to differentiate the false stop peaks from the real  
15 sequencing peaks by calculating the mass difference of the neighboring peaks and crs comparing the four spectra. Additionally, mt primers generated detectable data through the region of the downstream primer binding site thereby covering the false stop region.

Using optimized procedures of amplification, sequencing, and  
20 conditioning, exons 5-8 of the p53 gene were successfully sequenced. Correct wildtype sequence data were obtained from all exons with a mass resolution about 300 to 800 over the entire mass range. The overall mass accuracy is 0.05% or better. The average amount of each sequencing fragment loaded on the MS sample holder is estimated to be  
25 50 fmol or less.

This example demonstrates the feasibility of sequencing exons of a human gene by MALDI-TOF MS. Compare to gel-based automated fluorescent DNA sequencing, the read lengths are shorter. Microchip technology can be incorporated to provide for parallel processing.

Sequencing products generated in the microtiter plate can be directly transferred to a microchip which serves as a launching pad for MALDI-TOF MS analysis. Robot-driven serial and parallel nanoliter dispensing tools are being used to produce 100-1000 element DNA arrays on <1" square chips with flat or geometrically altered (e.g., with wells) surfaces for rapid mass spectrometric analysis.

Figure 94 shows an MS spectrum obtained on a chip where the sample was transferred from a microtiter plate by a pintool. The estimated amount of each termination product loaded is 5 fmol or less which is in the range of amounts used in conventional Sanger sequencing with radiolabeled or fluorescent detection (0.5-1 fmol per fragment). The low volume MALDI sample deposition has the advantages of miniaturization (reduced reagent cts), enhanced reproducibility and automated signal acquisition.

**15** **EXAMPLE 26**

## Direct detection of synthetic and biologically generated double-stranded DNA by MALDI-TOF MS

## Introduction

Typically, matrix-associated laser desorption/ionization (Karas, et. al., (1989) *Int. J. Mass Spectrom. Ion Processes*, 92, 231) time-of-flight mass spectrometry (MALDI-TOF MS) of DNA molecules which are double stranded (ds) in solution yields molecular ions representative of the two single stranded components (Tang, et al. (1994) Rapid Commun. Mass Spectrom. 8:183; Tang, et al. (1995) Nucleic Acids Res. 23:3126; Benner, et al. (1995) Rapid Commun. Mass Spectrom. 9:537; Liu, et al. (1995) Anal. Chem. 67:3482; Siegert et al. (1996) Anal. Biochem. 243:55; and Doktycz, et al. (1995) Anal. Biochem. 230:205); this has been observed in several reports dealing with biologically generated DNA from a polymerase chain reaction (PCR) amplification (Tang, et al. (1994)

-206-

Rapid Commun. Mass Spectrom. 8:183; Liu, *et al.* (1995) Anal. Chem. 67:3482; Siegert *et al.* (1996) Anal. Biochem. 243:55; and Doktycz, *et al.* (1995) Anal. Biochem. 230:205). It is not clear whether the double strand is destabilized because of the decreased pH in the matrix environment or because of absorbance by the duplex during desorption/ionization/acceleration of an energy sufficient to overcome the attractive van der Waals and "stacking" stabilization forces (Cantor and Shimmel, Biophysical Chemistry Part I: The conformation of Biomolecules, W.H. Freeman, New York, (1980), 176). When analyte is present at high concentrations formation of non-specific gas-phase DNA multimers is, as with proteins (Karas, *et. al.*, (1989) Int. J. Mass Spectrom, Ion Processes 92:231), common; however, Lecchi and Pannell (Lecchi *et al.* (1995) J. Am. Soc. Mass Spectrom. 6:972) have provided strong evidence for *specific* Watson Crick (WC) base pairing being maintained in the gas phase. They detected these specific dimers when using 6-aza-2-thiothymine as a matrix, but did not observe them with 3-hydroxypicclenic acid (3-HPA) or 2,4,6- hydroxyacetophenone matrix. As described below, by using a low acceleration voltage of the ions and preparing samples for MALDI analysis at reduced temperatures, routine detection of dsDNA is possible.

#### MATERIALS AND METHODS

**Synthetic DNA.** Oligonucleotides were synthesized (Sinha, *et al.* (1984) Nucleic Acids Res., 12, 4539) on a Perspective Expedite DNA synthesizer and reverse phase HPLC purified in-house. Sequences were:

25 50-mer (15337 Da): 5'-TTG CGT ACA CAC TGG CCG TCG TTT TAC AAC GTC GTG ACT GGG AAA ACC CT-3' (SEQ ID NO. 109); 27-mer<sub>c</sub> (complementary, 8343 Da): 5'-GTA AAA CGA CGG CCA GTG TGT ACG CAA-3' (SEQ ID NO. 110); 27-mer<sub>nc</sub> (non-complementary, 8293 Da): 5'-TAC TGG AAG GCG ATC TCA GCA ATC AGC-3' (SEQ ID NO. 111).

-207-

100 $\mu$ M stock solutions were diluted to 20, 10, 5, and 2.5  $\mu$ M using 18Mohm/cm H<sub>2</sub>O. 2 $\mu$ L each of equimolar solutions of the 50-mer and either 27-mer<sub>c</sub> or 27-mer<sub>nc</sub> were mixed and allowed to anneal at room temperature for 10 minutes. 0.5 $\mu$ L of these mixtures were mixed directly  
5 on a sample target with 1  $\mu$ L matrix (0.7 M 3-HPA, 0.07 M ammonium citrate in 50% acetonitrile) and allowed to air dry.

**Biological DNA.** Enzymatic digestion of human genomic DNA from leukocytes was performed. PCR primers (forward, 5'-GGC ACG GCT GTC CAA GGA G-3' (SEQ ID NO. 112)); reverse, 5'-AGG CCG CGC TCG  
10 GCG CCC TC-3' (SEQ ID NO. 113) to amplify a portion of exon 4 of the apolipoprotein E gene were delineated from the published sequence (Das et al., (1985) *J. Biol. Chem.*, 260 6240). Taq polymerase and 10x buffer were purchased from Boehringer-Mannheim (Germany) and dNTPs from Pharmacia (Freiburg, Germany). The total reaction volume was 50  
15  $\mu$ L including 20 pmol of each primer and 10% DMSO (dimethylsulfoxide, Sigma) with approximately 200 ng of genomic DNA used as template. Solutions were heated to 80°C before the addition of IU polymerase; PCR conditions were: 2 min at 94°C, followed by 40 cycles of 30 sec at 94°C, 45 sec at 63°C, 30 sec at 72°C, and a final extension time of 2  
20 min at 72°C. While no quantitative data was collected to determine the final yield of amplified product, it is estimated that ~2pmol were available for the enzymatic digestion.

CfoI and RsaI and reaction buffer L were purchased from Boehringer-Mannheim. 20 $\mu$ L of amplified products were diluted with 15 $\mu$ L  
25 water and 4 $\mu$ L buffer L; after addition of 10 units of restriction enzymes the samples were incubated for 60 min at 37°C. For precipitation of digest products 5 $\mu$ L of 3M ammonium acetate (pH 6.5), (5 $\mu$ L glycogen (Braun, et al. (1997) *Clin. Chem.* 43:1151) (10mg/ml, Sigma), and 110 $\mu$ L absolute ethanol were added to 50 $\mu$ L of the analyte solutions and stored

-208-

for 1 hour at room temperature. After at 10 min centrifugation at 13,000 X g, the pellet was washed in 70% ethanol and resuspended in 1  $\mu$ l 18Mohm/cm H<sub>2</sub>O.

*Sample preparation and analysis by MALDI-TOF MS.* 0.35  $\mu$ l of resuspended DNA was mixed with 0.35-1.3  $\mu$ l matrix solution (0.7M 3-hydroxypicolinic acid (3-HPA), 0.07 M ammonium citrate in 1:1 H<sub>2</sub>O:CH<sub>3</sub>CN) (Wu, et al. (1993) Rapid Commun. Mass Spectrom. 7:142) on a stainless steel sample target disk and allowed to air dry preceding spectrum acquisition using a Thermo Bioanalysis Vision 2000 MALDI-TOF instrument operated in pitive ion reflectron mode with 5 and 20 kV on the target and conversion dynode, respectively. Theoretical average molecular masses ( $M_r(\text{calc})$ ) of the fragments were calculated from atomic compositions; the mass of a proton (1.08 Da) was subtracted from raw data values in reporting experimental molecular masses ( $M_r(\text{exp})$ ) as neutral basis. External calibration generated from eight peaks (2000-18000 Da) was used for all spectra.

#### Results and Discussion

Figure 96A is a MALDI-TOF mass spectrum of a mixture of the synthetic 50-mer with (non-complementary) 27-mer<sub>nc</sub> (each 10  $\mu$ M, the highest final concentration used in this study); the laser power was adjusted to just above the threshold irradiation for ionization. The peaks at 8.30 and 15.34 kDa represent singly charged ions derived from the 27- and 50-mer single strands, respectively. Poorly resolved low intensity signals at -16.6 and -30.7 kDa represent homodimers of 27- and 50-mer, respectively; that at 23.6 kDa is consistent with a heterodimer containing one 27-mer and one 50-mer strand. Thus low intensity dimer ions representing all possible combinations from the two non-complementary oligonucleotides (27 + 27; 27 + 50; 50 + 50) were observed. Increasing the irradiance even to a point where depurination

-209-

peaks dominated the spectrum resulted in slightly higher intensities of these dimer peaks. Note that the hybridization was performed at room temperature and with a very low salt concentration, conditions at which non-specific hybridization may occur.

- 5        Figure 96 shows a MALDI-TOF spectrum of the same 50-mer mixed with (complementary) 27-mer; the final concentration of each oligonucleotide was again 10  $\mu$ M. Using the same laser power as in Figure 96A, intense signals were again observed at 88.34 and 15.34 Kda, consistent with single stranded 27- and 50-mer, respectively.
- 10      Homodimer peaks (27 + 27; 50 + 50) were barely apparent in the noise; however, singly (23.68 Kda) and doubly (11.84k Da) charged heterodimer (27 + 50) peaks were dominant. Although the 23.68 Kda dimer peak could be detected from all irradiated positions, its intensity relative to the monomer peaks varied slightly from spot-to-spot.
- 15      Repeating the experiment with individual oligonucleotide concentrations of 5, 2.5, and 1.25  $\mu$ M resulted in decreasing amounts of the 27-/50-mer Watson-Crick dimer peak relative to the 27- and 50-mer single stranded peaks. At the lowest concentrations, the observation of dimer was "crystal-dependent", that is, irradiation of some crystals produced
- 20      significant 27-/50-mer dimer signal, while other crystals reproducibly yielded very little or none. This indicates that the incorporation of dsdna into the matrix crystals or the effectiveness of retaining this interaction through the ionization/desorption process is dependent upon the microscopic properties of the crystals, and/or that there exist steep
- 25      concentration gradients of the duplex throughout the sample.

Thus the Figure 96 spectra provide strong evidence that specific WC base paired dsdna can be observed using gentle laser conditions with high concentrations of oligonucleotides in this mass range, the first report of this using a 3-HPA matrix. The study was extended to a

-210-

complex mixture of dsdna derived from an enzymatic digest (RsaI/CfoI) of a region of exon 4 of the apolipoprotein E gene (Das et. al., (1985) *J. Biol. Chem.*, 260 6240); expected fragment masses are given in

Table IX.

5

**Table IX**  
**CfoI/RsaI Digestion Products from ApoE gene exon 4<sup>2</sup>**

10

15

bases <sup>b</sup>		ssDNA	(Da)	dsdna (Da)
(+)	(-)	(+)	(-)	
11	13	3428	4025	7453
16		5004	4924	9928
18		5412	5750	11162
17	19	5283	5880	11163
19		5999	5781	11780
24	22	7510	6745	14225
31	29	9628	9185	18813
36	38	11279	11627	22906
48		14845	14858	29703
55	53	17175	16240	33415

20 <sup>a</sup>ε3 allele has no 17/19 or 19/19 pairs; ε4 allele contains no 36/38 pair.

<sup>b</sup>(+) sense strand, (-) antisense strand

After the digestion step, the samples were purified and concentrated by ethanol precipitation and resuspended in 1μL H<sub>2</sub>O before mixing them at room temperature with matrix on the sample target. Nearly 20 peaks

25 ranging in mass from 3.4-17.2 Kda were resolved in the products'

MALDI spectrum (Figure 97A), all consistent with denatured single stranded components of the double strand (Table IX). Many such analyses of similar biological products over a period of months also yielded spectra with negligible dsdna, consistent with previous reports

30 (Tang, et al. (1994) *Rapid Commun. Mass Spectrom.* 8:183; Liu, et al.



-211-

(1995) Anal. Chem. 67:3482; Siegert *et al.* (1996) Anal. Biochem. 243:55; and Doktycz, *et al.* (1995) Anal. Biochem. 230:205); contrarily, intact double strands were observed under similar conditions for the synthetic DNA (Figure 96A). It is difficult to estimate the strand concentration available after the biological reactions, but presumably that it was far lower than that at which dimerization of synthetic samples occurred. Furthermore, maintaining specific hybrids within the two-component synthetic mixture may be kinetically favored relative to the far more complex mixture of 20 single-stranded DNA components from the digest.

The effect of reduced temperature on maintaining dsDNA was tested. An aliquot of the digested DNA solution, the matrix, pipette, pipette tips, and the stainless steel sample target were stored in a 4°C "cold room" for 15 minutes; as with normal preparations matrix, and then analyte, were spotted on the target and allowed to co-crystallize while air drying. Crystallization for mixtures of 300 nL 3HPA (50% acetonitrile) with 300 nL analyte required ~ 1 minute at room temperature but ~ 15 minutes at the reduced temperature. Sample spots prepared in the cold room environment typically contained a high proportion of large transparent crystals.

MALDI-TOF analysis of an ApoE digest aliquot prepared at reduced temperature produced the Figure 97B spectrum. While the low mass range appeared qualitatively similar to Figure 97A, dramatic differences above 8 kDa were observed. Only signals consistent with single strands (Table IX) were observed in Figure 97A, but the Figure 97B cold room prepared samples did not yield signals for the same masses except below 8 kDa. Even more striking were the additional high mass peaks in Figure 97B; clearly these represent dimer peaks containing lower mass components. As was done with the synthetic DNA, it was important to

-212-

determine whether these represent non-specific heterodimers, specific WC heterodimers, or nonspecific homodimers. Consider first the 33.35 kDa fragment. Ignoring the unlikely possibility that the high mass fragment represents a trimer or higher multimer, as a dimer it must only

5 contain the highest mass ssDNA components, i.e., the > 16 kDa. Homodimerization of the 15.24 and 17.18 kDa fragments would result in 32.49 and 34.35 kDa peaks, respectively; corresponding mass errors for these incorrect assignments relative to the observed 33.35 kDa would be -2.6% and +3.0% respectively. A far better match is achieved if this

10 peak originates from a heterodimer of the two highest mass single stranded fragments; their summed mass ( $16.24 + 17.18 = 33.42$  kDa) differed by 0.2% from the observed dimer mass 33.35 kDa, an acceptable mass error for MALDI-TOF analysis of large DNA fragments using external calibration. Likewise, the 29.66 kDa fragment was

15 measured only 0.13% lower than the 29.70 Da expected for a heterodimer of 48-mers; the sum of no other possible homodimers or heterodimers were within a reasonable range of this mass. Similar arguments could be made for the 22.89 and 18.83 kDa fragments, representing 36-/38-mer and 31-/29-heterodimers, respectively; the

20 signal at 14.86 kDa is consistent with singly charged single stranded and doubly charged double-stranded 48-mer. The agreement of the Figure 97B masses above 15 kDa with the of dsDNA expected from this digest and the absence of homodimers and non-specific heterodimers at random masses indicated that the base pairings were indeed highly specific and

25 provided further evidence that gas-phase WC interactions may be retained in MALDI-generated ions.

Figure 98 shows a MALDI-TOF spectrum of an  $\epsilon 4$  allele, which, unlike the  $\epsilon 3$ , was expected to yield no 36-/38-mer pair upon CfoI/RsaI digestion. The  $\epsilon 3$  and  $\epsilon 4$  mass spectra were similar except that

-213-

abundant 22.89 kDa fragment in Figure 97B was not present in Figure 98; with this information alone (Table IX)  $\epsilon 3$  and  $\epsilon 4$  alleles were easily distinguished, thereby demonstrating the genotyping by direct measurement of dsDNA by MALDI-TOF MS. Similarly dsDNA could be ionized, transferred to the gas phase, and detected by MALDI-TOF MS. The acceleration voltage typically employed on our instrument was only 5kV corresponding to 1.5kV/mm up to -2 mm from the sample target, with the electric field strength decreasing rapidly with distance from the sample target. Most previous work used at least 20kV acceleration (Lecchi *et al.* (1995) J. Am. Soc. Mass Spectrom. 6:972); in one exception a 27-mer dsDNA was detected using a frozen matrix solution and 100 V acceleration (Nelson, *et al.* (1990) Rapid Commun. Mass Spectrom. 4:348). Without being bound by any theory MALDI-induced "denaturation" of dsDNA may be due to gas-phase collisional activation that disrupts the WC pairing when high acceleration fields are employed, analogous to the denaturation presumed to be a first step in the fragmentation used for sequencing the single stranded components of dsDNA using electrospray ionization (McLafferty *et al.* (1996) Int. J. Mass Spectrom., Ion Processes). It appears that the high salt concentrations (typically >10mM NaCl or KCl) required to stabilize WC paired dsDNA in solution are unsuitable for MALDI analysis (Nordhoff *et al.* (1993) Nucleic Acids Res. 21:3347); reducing the concentration of such non-volatile cations is necessary to avoid cation-adducted MALDI signals, but destabilizes the double strands in solution. The low pH conditions of the matrix environment should also destabilize the duplex. As shown in Figures 97B and 98, storing and preparing even low concentrations of the biological samples at reduced temperature at least in part offset these denaturing effects, especially for longer strands where melting temperatures are higher due to a more extensive hydrogen

-214-

bonding network. The conditions used here are recognized to be very non-stringent annealing conditions.

The low mass tails on high mass dsDNA peaks (e.g., Figure 97B, 232 kDa) are consistent with depurination generated to a higher extent than the sum of depurination from each of the single strands combined. Although depurination in solution is an acid-catalyzed reaction, the weakly acidic conditions in the 3-HPA matrix do not induce significant depurination; molecular ion signals from a mixed-base 50-mer measured with De-MALDI-TOF had only minor contributions from depurination peaks (Juhaz, *et al.* (1996) Anal. Chem. 68:941). Depurination from the single stranded components of the gas-phase dsDNA is observed even though these bases are expected to be hydrogen bonded to the complementary base of the accompanying strand, implying that covalent bonds are being broken before the strand is denatured.

15

#### EXAMPLE 27

##### Efficiency and Specificity Assay for Base-Specific Ribonucleases

Aliquots sampled at regular time intervals during digestion of selected synthetic 20 to 25 mers were analyzed by mass spectrometry. Three of the RNAses were found to be efficient and specific. These include: the G-specific  $T_1$ , the A-specific  $U_2$  and the A/U-specific PhyM. The ribonucleases presumed to be C-specific were found to be less reliable, e.g., did not cleave at every C or also cleaved at U in an unpredictable manner. The three promising RNAses all yielded cleavage at all of the predicted positions and a complete sequence coverage was obtained. In addition, the presence of cleavage products containing one or several uncleaved positions (short incubation times), allowed alignment of the cleavage products. An example of the MALDI-spectrum of an aliquot sampled after  $T_1$  digest of a synthetic 20-mer [SEQ ID NO:114] RNA is shown in Figure 100.

-215-

**EXAMPLE 28****Immobilization of amplified DNA targets to silicon wafers****Silicon surface preparation**

Silicon wafers were washed with ethanol, flamed over bunsen  
5 burner, and immersed in an anhydrous solution of 25% (by volume) 3-aminopropyltriethoxysilane in toluene for 3 hours. The silane solution was then removed, and the wafers were washed three times with toluene and three times with dimethyl sulfoxide (DMSO). The wafers were then incubated in a 10mM anhydrous solution of N-succinimidyl (4-  
10 iodoacetyl) aminobenzoate (SIAB) (Pierce Chemical, Rockford, IL) in anhydrous DMSO. Following the reaction, the SIAB solution was removed, and the wafers were washed three times with DMSO. In all cases, the iodoacetamido-functionalized wafers were used immediately to minimize hydrolysis of the labile iodoacetamido-functionality.  
15 Additionally, all further wafer manipulations were performed in the dark since the iodoacetamido-functionality is light sensitive.

**Immobilization of amplified thiol-containing nucleic acids**

The SIAB-conjugated silicon wafers were used to analyze specific free thiol-containing DNA fragments of a particular amplified DNA target  
20 sequence. A 23-mer oligodeoxynucleotide containing a 5'-disulfide linkage [purchased from Operon Technologies; SEQ ID NO: 117] that is complementary to the 3'-region of a 112 bp human genomic DNA template [Genebank Acc. No.: Z52259; SEQ ID NO: 118] was used as a primer in conjunction with a commercially available 49-mer primer, which  
25 is complementary to a portion of the 5'-end of the genomic DNA [purchased from Operon Technologies; SEQ ID NO: 119], in PCR reactions to amplify a 135 bp DNA product containing a 5'-disulfide linkage attached to only one strand of the DNA duplex [SEQ ID NO: 120].

-216-

The PCR amplification reactions were performed using the Amplitaq GoldKit [Perkin Elmer Catalog No. N808-0249]. Briefly, 200 ng 112 bp human genomic DNA template was incubated with 10  $\mu$ M of 23-mer primer and 8  $\mu$ M of commercially available 49-mer primer, 10 mM dNTPs, 1 unit of Amplitaq Gold DNA polymerase in the buffer provided by the manufacturer and PCR was performed in a thermocycler.

The 5'-disulfide bond of the resulting amplified product was fully reduced using 10 mM tris-(2-carboxyethyl) phosphine (TCEP) (Pierce Chemical, Rockford, IL) to generate a free 5'-thiol group. Disulfide reduction of the modified oligonucleotide was monitored by observing a shift in retention time on reverse-phase FPLC. It was determined that after five hours in the presence of 10 mM TCEP, the disulfide was fully reduced to a free thiol. Immediately following disulfide cleavage, the modified oligonucleotide was incubated with the iodacetamido-functionalized wafers and conjugated to the surface of the silicon wafer through the SIAB linker. To ensure complete thiol deprotonation, the coupling reaction was performed at pH 8.0. Using 10mM TCEP to cleave the disulfide and the other reaction conditions described above, it was possible to reproducibly yield a surface density of 250 fmol per square mm of surface.

#### Hybridization and MALDI-TOF Mass spectrometry

The silicon wafer conjugated with the 135 bp thiol-containing DNA was incubated with a complementary 12-mer oligonucleotide [SEQ ID NO: 121] and specifically hybridized DNA fragments were detected using MALDI-TOF MS analysis. The mass spectrum revealed a signal with an observed experimental mass-to-charge ratio of 3618.33; the theoretical mass-to-charge ratio of the 12-mer oligomer sequence is 3622.4 Da.

Thus, specific DNA target molecule that contain a 5'-disulfide linkage can be amplified. The molecules are immobilized at a high

-217-

density on a SIAB-derivatized silicon wafer using the methods described herein and specific complementary oligonucleotides may be hybridized to these target molecules and detected using MALDI-TOF MS analysis.

#### EXAMPLE 29

##### 5 Use of High Density Nucleic Acid Immobilization to Generate Nucleic Acid Arrays

Employing the high density attachment procedure described in EXAMPLE 28, an array of DNA oligomers amenable to MALDI-TOF mass spectrometry analysis was created on a silicon wafer having a plurality of  
10 locations, e.g., depressions or patches, on its surface. To generate the array, a free thiol-containing oligonucleotide primer was immobilized only at the selected locations of the wafer [e.g., see EXAMPLE 28]. The each location of the array contained one of three different oligomers. To demonstrate that the different immobilized oligomers could be separately  
15 detected and distinguished, three distinct oligonucleotides of differing lengths that are complementary to one of the three oligomers were hybridized to the array on the wafer and analyzed by MALDI-TOF mass spectrometry.

##### Oligodeoxynucleotides

20 Three sets of complementary oligodeoxynucleotide pairs were synthesized in which one member of the complementary oligonucleotide pair contains a 3'- or 5'-disulfide linkage [purchased from Operon Technologies or Oligos, Etc.]. For example, Oligomer 1 [d(CTGATGCGTCGGATCATCTTTTTT-SS); SEQ ID NO: 122] contains a  
25 3'-disulfide linkage whereas Oligomer 2 [d(SS-CCTCTTGGGAAGTGTAGTATT); a 5'-disulfide derivative of SEQ ID NO: 117] and Oligomer 3 [d(SS-GAATTCGAGCTCGGTACCCGG); a 5'-disulfide derivative of SEQ ID NO: 115] each contain a 5'-disulfide linkage .

-218-

The oligonucleotides complementary to Oligomers 1-3 were designed to be of different lengths that are easily resolvable from one another during MALDI-TOF MS analysis. For example, a 23-mer oligonucleotide [SEQ ID NO: 123] was synthesized complementary to a portion of Oligomer 1, a 12-mer oligonucleotide [SEQ ID NO: 121] was synthesized complementary to a portion of Oligomer 2 and a 21-mer [SEQ ID NO: 116] was synthesized complementary to a portion of Oligomer 3. In addition, a fourth 29-mer oligonucleotide [SEQ ID NO: 124] was synthesized that lacks complementarity to any of the three oligomers. This fourth oligonucleotide was used as a negative control.

#### **Silicon surface chemistry and DNA immobilization**

##### **(a) 4 x 4 (16-location) array**

A 2 X 2 cm<sup>2</sup> silicon wafer having 256 individual depressions or wells in the form of a 16 X 16 well array was purchased from a commercial supplier [Accelerator Technology Corp., College Station, Texas]. The wells were 800 X 800  $\mu\text{m}^2$ , 120  $\mu\text{m}$  deep, on a 1.125 pitch. The silicon wafer was reacted with 3-aminopropyltriethoxysilane to produce a uniform layer of primary amines on the surface and then exposed to the heterobifunctional crosslinker SIAB resulting in iodoacetamido functionalities on the surface [e.g., see EXAMPLE 28].

To prepare the oligomers for coupling to the various locations of the silicon array, the disulfide bond of each oligomer was fully reduced using 10 mM TCEP as depicted in EXAMPLE 28, and the DNA resuspended at a final concentration of 10  $\mu\text{M}$  in a solution of 100 mM phosphate buffer, pH 8.0. Immediately following disulfide bond reduction, the free-thiol group of the oligomer was coupled to the iodoacetamido functionality at 16 locations on the wafer using the probe coupling conditions essentially as described above in EXAMPLE 28. To accomplish the separate coupling at 16 distinct locations of the wafer,



-219-

the entire surface of the wafer was not flushed with an oligonucleotide solution but, instead, an ~30-nl aliquot of a predetermined modified oligomer was added in parallel to each of 16 locations (i.e., depressions) of the 256 wells on the wafer to create a 4 x 4 array of immobilized DNA using a robotic pintool.

The robotic pintool consists of 16 probes housed in a probe block and mounted on an X Y, Z robotic stage. The robotic stage was a gantry system which enables the placement of sample trays below the arms of the robot. The gantry unit itself is composed of X and Y arms which  
10 move 250 and 400 mm, respectively, guided by brushless linear servo motors with positional feedback provided by linear optical encoders. A lead screw driven Z axis (50 mm vertical travel) is mounted to the xy axis slide of the gantry unit and is controlled by an in-line rotary servo motor with positional feedback by a motor-mounted rotary optical encoder. The  
15 work area of the system is equipped with a slide-out tooling plate that holds five microtiter plates (most often, 2 plates of wash solution and 3 plates of sample for a maximum of 1152 different oligonucleotide solutions) and up to ten 20x20 mm wafers. The wafers are placed precisely in the plate against two banking pins and held secure by  
20 vacuum. The entire system is enclosed in plexi-glass housing for safety and mounted onto a steel support frame for thermal and vibrational damping. Motion control is accomplished by employing a commercial motion controller which was a 3-axis servo controller and is integrated to a computer; programming code for specific applications is written as  
25 needed.

To create the DNA array, a pintool with assemblies that have solid pin elements was dipped into 16 wells of a multi-well DNA source plate containing solutions of Oligomers 1-3 to wet the distal ends of the pins, the robotic assembly moves the pin assembly to the silicon wafer, and

-220-

the sample spotted by surface contact. Thus, one of modified Oligomers 1-3 was covalently immobilized to each of 16 separate wells of the 256 wells on the silicon wafer thereby creating a 4 x 4 array of immobilized DNA.

- 5 In carrying out the hybridization reaction, the three complementary oligonucleotides and the negative control oligonucleotide were mixed at a final concentration of 10  $\mu$ M for each oligonucleotide in 1 ml of TE buffer [10 mM Tris-HCl, pH 8.0, 1 mM EDTA] supplemented with 1 M NaCl, and the solution was heated at 65°C for 10 min. Immediately thereafter,
- 10 the entire surface of the silicon wafer was flushed with 800  $\mu$ l of the heated oligonucleotide solution. The complementary oligonucleotides were annealed to the immobilized oligomers by incubating the silicon array at ambient temperature for 1 hr, followed by incubation at 4°C for at least 10 min. Alternatively, the oligonucleotide solution can be added
- 15 to the wafer which is then heated and allowed to cool for hybridization.

The hybridized array was then washed with a solution of 50 mM ammonium citrate buffer for cation exchange to remove sodium and potassium ions on the DNA backbone (Pieles *et al.*, (1993) Nucl. Acids Res. 21:3191-3196). A 6- $\mu$ l aliquot of a matrix solution of 3-

20 hydroxypicolinic acid [0.7 M 3-hydroxypicolinic acid-10 % ammonium citrate in 50 % acetonitrile; see Wu *et al.* Rapid Commun. Mass Spectrom. 7:142-146 (1993)] was added in series to each location of the array using a robotic piezoelectric serial dispenser (*i.e.*, a piezoelectric pipette system).

- 25 The piezoelectric pipette system is built on a system purchased from Microdrop GmbH, Norderstedt Germany and contains a piezoelectric element driver which sends a pulsed signal to a piezoelectric element bonded to and surrounding a glass capillary which holds the solution to be dispensed; a pressure transducer to load (by negative pressure) or

-221-

empty (by positive pressure) the capillary; a robotic xyz stage and robot driver to maneuver the capillary for loading, unloading, dispensing, and cleaning, a stroboscope and driver pulsed at the frequency of the piezo element to enable viewing of 'suspended' droplet characteristics;

- 5 separate stages for source and designation plates or sample targets (i.e. Si chip); a camera mounted to the robotic arm to view loading to designation plate; and a data station which controls the pressure unit, xyz robot, and piezoelectric driver.

- The 3-HPA solution was allowed to dry at ambient temperature  
10 and thereafter a 6- $\mu$ l aliquot of water was added to each location using the piezoelectric pipette to resuspend the dried matrix-DNA complex, such that upon drying at ambient temperature the matrix-DNA complex forms a uniform crystalline surface on the bottom surface of each location.

15 **MALDI-TOF MS analysis**

- The MALDI-TOF MS analysis was performed in series on each of the 16 locations of the hybridization array illustrated in Figure 6 essentially as described in EXAMPLE 28. The resulting mass spectrum of oligonucleotides that specifically hybridized to each of the 16 locations of  
20 the DNA hybridization revealed a specific signal at each location representative of observed experimental mass-to-charge ratio corresponding to the specific complementary nucleotide sequence.

- For example, in the locations that have only Oligomer 1 conjugated thereto, the mass spectrum revealed a predominate signal with an  
25 observed experimental mass-to-charge ratio of 7072.4 approximately equal to that of the 23-mer; the theoretical mass-to-charge ratio of the 23-mer is 7072.6 Da. Similarly, specific hybridization of the 12-mer oligonucleotide to the array, observed experimental mass-to-charge ratio of 3618.33 Da (theoretical 3622.4 Da), was detected only at those

-222-

locations conjugated with Oligomer 2 whereas specific hybridization of MJM6 (observed experimental mass-to-charge ratio of 6415.4) was detected only at those locations of the array conjugated with Oligomer 3 [theoretical 6407.2 Da].

- 5        None of the locations of the array revealed a signal that corresponds to the negative control 29-mer oligonucleotide (theoretical mass-to-charge ratio of 8974.8) indicating that specific target DNA molecules can be hybridized to oligomers covalently immobilized to specific locations on the surface of the silicon array and a plurality of
- 10    hybridization assays may be individually monitored using MALDI-TOF MS analysis.

(b)    8 x 8 (64-location) array

- A 2 X 2 cm<sup>2</sup> silicon wafer having 256 individual depressions or wells that form a 16 X 16 array of wells was purchased from a
- 15    commercial supplier [Accelerator Technology Corp., College Station, Texas]. The wells were 800 X 800  $\mu\text{m}^2$ , 120  $\mu\text{m}$  deep, on a 1.125 pitch. The silicon wafer was reacted with 3-aminopropyltriethoxysilane to produce a uniform layer of primary amines on the surface and then exposed to the heterobifunctional crosslinker SIAB resulting in
- 20    iodoacetamido functionalities on the surface as described above.

- To make an array of 64 elements, a pintool was used following the procedures described above. The pintool was dipped into 16 wells of a 384 well DNA source plate containing solutions of Oligomers 1-3, moved to the silicon wafer, and the sample spotted by surface contact. Next,
- 25    the tool was dipped in washing solution, then dipped into the same 16 wells of the source plate, and spotted onto the target 2.25mm offset from the initial set of 16 spots; the entire cycle was repeated to make a 2x2 array from each pin to produce an 8x8 array of spots (2x2 elements/pin X 16 pins = 64 total elements spotted).

-223-

Oligomers 1-3 immobilized to the 64 locations were hybridized to complementary oligonucleotides and analyzed by MALDI-TOF MS analysis. As observed for the 16-location array, specific hybridization of the complementary oligonucleotide to each of the immobilized thiol-  
5 containing oligomers was observed in each of the locations of the DNA array.

### EXAMPLE 30

**Extension of hybridized DNA primers bound to DNA templates immobilized on a silicon wafer**

10 The SIAB-derivatized silicon wafers can also be employed for primer extension reactions of the immobilized DNA template using the procedures essentially described in EXAMPLE 7.

A 27-mer oligonucleotide [SEQ ID NO: 125] containing a 3'-free thiol group was coupled to a SIAB-derivatized silicon wafer as described  
15 above, for example, in EXAMPLE 28. A 12-mer oligonucleotide primer [SEQ ID NO: 126] was hybridized to the immobilized oligonucleotide and the primer was extended using a commercially available kit [e.g., Sequenase or ThermoSequenase, U.S. Biochemical Corp]. The addition of Sequenase DNA polymerase or ThermoSequenase DNA polymerase in  
20 the presence of three deoxyribonucleoside triphosphates (dNTPs; dATP, dGTP, dCTP) and dideoxyribonucleoside thymidine triphosphate (ddTTP) in buffer according to the instructions provided by the manufacturer resulted in a 3-base extension of the 12-mer primer while still bound to the silicon wafer. The wafer was then analyzed by MALDI-TOF mass  
25 spectrometry as described above. The mass spectrum results clearly distinguish the 15-mer [SEQ ID NO: 127] from the original unextended 12-mer thus indicating that specific extension can be performed on the surface of a silicon wafer and detected using MALDI-TOF MS analysis.

-224-

**EXAMPLE 31**

**Effect of linker length on polymerase extension of hybridized DNA primers bound to DNA templates immobilized on a silicon wafer**

The effect of the distance between the SIAB-conjugated silicon surface and the duplex DNA formed by hybridization of the target DNA to the immobilized oligomer template was investigated, as well as choice of enzyme.

Two SIAB-derivatized silicon wafers were conjugated to the 3'-end of two free thiol-containing oligonucleotides of identical DNA sequence except for a 3-base poly dT spacer sequence incorporated at the 3'-end:

CTGATGCGTC GGATCATCTT TTTT SEQ ID No. 122

CTGATGCGTC GGATCATCTT TTTTTT SEQ ID No. 125.

These oligonucleotides were synthesized and each was separately immobilized to the surface of a silicon wafer through the SIAB cross-linker [e.g., see EXAMPLE 28]. Each wafer was incubated with a 12-mer oligonucleotide:

AAAAAAGATG AT SEQ ID No. 126

GATGATCCGA CG SEQ ID No. 128

GATCCGACGC AT SEQ ID No. 129,

which is complementary to portions of the nucleotide sequences common to both of the oligonucleotides, by denaturing at 75 °C and slow cooling the silicon wafer. The wafers were then analyzed by MALDI-TOF mass spectrometry as described above.

As described in EXAMPLE 30 above, a 3-base specific extension of the bound 12-mer oligonucleotide was observed using the oligomer primer where there is a 9-base spacer between the duplex and the surface [SEQ ID NO: 125]. Similar results were observed when the DNA spacer lengths between the SIAB moiety and the DNA duplex were 0, 3, 6 and 12. In addition, the extension reaction may be performed using a

variety of DNA polymerases, such as Sequenase and Thermo Sequenase (US Biochemical). Thus, the SIAB linker may be directly coupled to the DNA template or may include a linker sequence without effecting primer extension of the hybridized DNA.

**5** **EXAMPLE 32**

## Spectrochip mutant detection in ApoE gene

This example describes the hybridization of an immobilized template, primer extension and mass spectrometry for detection of the wildtype and mutant Apolipoprotein E gene for diagnostic purposes.

- 10 This example demonstrates that immobilized DNA molecules containing a specific sequence can be detected and distinguished using primer extension of unlabeled allele specific primers and analysis of the extension products using mass spectrometry.**

A 50 base synthetic DNA template complementary to the coding  
15 sequence of allele 3 of the wildtype apolipoprotein E gene:

5'- GCCTGGTACACTGCCAGGCCTTCTGCAGGTCATCGGCATCGCGGAGGAG -3'  
[SEQ ID NO: 280]

or complement to the mutant apolipoprotein E gene carrying a G → A transition at codon 158:

- 20** 5'-GCCTGGTACACTGCCAGGCACTTCTGCAGGTCATCGGCATCGCGGAGGAG-3'  
[SEQ ID NO: 281]

containing a 3'-free thiol group was coupled to separate SIAB-derivatized silicon wafers as described in Example 28.

**A 21-mer oligonucleotide primer:**

- 25 5'-GAT GCC GAT GAC CTG CAG AAG-3' [SEQ ID NO: 282] was hybridized to each of the immobilized templates and the primer was extended using a commercially available kit [e.g., Sequenase or Thermosequenase, U.S. Biochemical Corp]. The addition of Sequenase DNA polymerase or Thermosequenase DNA polymerase in the presence  
30 of three deoxyribonucleoside triphosphates (dNTPs; dATP, dGTP, dTTP)

-226-

and dideoxyribonucleoside cytosine triphosphate (ddCTP) in buffer according to the instructions provided by the manufacturer resulted in a single base extension of the 21-mer primer bound to the immobilized template encoding the wildtype apolipoprotein E gene and a three base extension of the 21-mer primer bound to the immobilized template encoding the mutant form of apolipoprotein E gene.

The wafers were analyzed by mass-spectrometry as described herein. The wildtype apolipoprotein E sequence results in a mass spectrum that distinguishes the primer with a single base extension (22-mer) with a mass to charge ratio of 6771.17 Da (the theoretical mass to charge ratio is 6753.5 Da) from the original 21-mer primer with a mass to charge ratio of 6499.64 Da. The mutant apolipoprotein E sequence results in a mass spectrum that distinguishes the primer with a three base extension (24-mer) with a mass to charge ratio of 7386.9 (the theoretical mass charge is 7386.9) from the original 21-mer primer with a mass to charge ratio of 6499.64 Da.

### EXAMPLE 33

#### Detection of Double-Stranded Nucleic Acid Molecules via Strand Displacement and Hybridization to an Immobilized Complementary Nucleic Acid

This example describes immobilization of a 24-mer primer and the specific hybridization of one strand of a duplex DNA molecule, thereby permitting amplification of a selected target molecule in solution phase and permitting detection of the double stranded molecule. This method is useful for detecting single base changes, and, particularly for screening genomic libraries of double-stranded fragments.

A 24-mer DNA primer CTGATGCGTC GGATCATCTT TTTT SEQ ID No. 122, containing a 3'-free thiol group was coupled to a SIAB-derivatized silicon wafer as described in Example 29.

An 18-mer synthetic oligonucleotide:



-227-

5'-CTGATGCGTCGGATCATC-3' [SEQ ID NO: 286] was premixed with a 12-mer 5'-GATGATCCGACG-3' [SEQ ID NO: 285] that has a sequence that is complementary to 12 base portion of the 18-mer oligonucleotide. The oligonucleotide mix was heated to 75°C and  
5 cooled slowly to room temperature to facilitate the formation of a duplex molecule:

5'-CTGATGCGTCGGATCATC-3' [SEQ ID NO: 286]  
3'-GCAGCCTAGTAG-5' [SEQ ID NO: 287].

The specific hybridization of the 12-mer strand of the duplex  
10 molecule to the immobilized 24-mer primer was carried out by mixing 1 μM of the duplex molecule using the hybridization conditions described in Example 30.

The wafers were analyzed by mass spectrometry as described above. Specific hybridization was detected in a mass spectrum of the  
15 12-mer with a mass to charge ratio of 3682.78 Da.

#### EXAMPLE 34

1-(2-Nitro-5-(3-O-4,4'-dimethoxytritylpropoxy)phenyl)-1-O-((2-cyanoethoxy)-diisopropylaminophosphino)ethane

##### A. 2-Nitro-5-(3-hydroxypropoxy)benzaldehyde

20 3-Bromo-1-propanol (3.34 g, 24 mmol) was refluxed in 80 ml of anhydrous acetonitrile with 5-hydroxy-2-nitrobenzaldehyde (3.34 g, 20 mmol), K<sub>2</sub>CO<sub>3</sub> (3.5 g), and KI (100 mg) overnight (15 h). The reaction mixture was cooled to room temperature and 150 ml of methylene chloride was added. The mixture was filtered and the solid residue was  
25 washed with methylene chloride. The combined organic solution was evaporated to dryness and redissolved in 100 ml methylene chloride. The resulted solution was washed with saturated NaCl solution and dried over sodium sulfate. 4.31 g (96%) of desired product was obtained after removal of the solvent in vacuo.

30 R<sub>f</sub> = 0.33 (dichloromethane/methanol, 95/5).

-228-

UV (methanol) maximum: 313, 240 (shoulder), 215 nm; minimum: 266 nm.

$^1\text{H}$  NMR ( $\text{DMSO}-d_6$ )  $\delta$  10.28 (s, 1H), 8.17 (d, 1H), 7.35 (d, 1H), 7.22 (s, 1H), 4.22 (t, 2H), 3.54 (t, 2H), 1.90 (m, 2H).

5  $^{13}\text{C}$  NMR ( $\text{DMSO}-d_6$ )  $\delta$  189.9, 153.0, 141.6, 134.3, 127.3, 118.4, 114.0, 66.2, 56.9, 31.7.

**B. 2-Nitro-5-(3-O-t-butyldimethylsilylpropoxy)benzaldehyde**

2-Nitro-5-(3-hydroxypropoxy)benzaldehyde (1 g, 4.44 mmol) was dissolved in 50 ml anhydrous acetonitrile. To this solution, it was added  
10 1 ml of triethylamine, 200 mg of imidazole, and 0.8 g (5.3 mmol) of tBDMSCl. The mixture was stirred at room temperature for 4 h. Methanol (1 ml) was added to stop the reaction. The solvent was removed in vacuo and the solid residue was redissolved in 100 ml methylene chloride. The resulted solution was washed with saturated  
15 sodium bicarbonate solution and then water. The organic phase was dried over sodium sulfate and the solvent was removed in vacuo. The crude mixture was subjected to a quick silica gel column with methylene chloride to yield 1.44 g (96%) of 2-nitro-5-(3-O-t-butyldimethylsilylpropoxy)benzaldehyde.  
20  $R_f = 0.67$  (hexane/ethyl acetate, 5/1).

UV (methanol), maximum: 317, 243, 215 nm; minimum: 235, 267 nm.

$^1\text{H}$  NMR ( $\text{DMSO}-d_6$ )  $\delta$  10.28 (s, 1H), 8.14 (d, 1H), 7.32 (d, 1H), 7.20 (s, 1H), 4.20 (t, 2H), 3.75 (t, 2H), 1.90 (m, 2H), 0.85 (s, 9H), 0.02 (s, 6H).

$^{13}\text{C}$  NMR ( $\text{DMSO}-d_6$ )  $\delta$  189.6, 162.7, 141.5, 134.0, 127.1, 118.2,  
25 113.8, 65.4, 58.5, 31.2, 25.5, -3.1, -5.7.

**C. 1-(2-Nitro-5-(3-O-t-butyldimethylsilylpropoxy)phenyl)ethanol**

High vacuum dried 2-nitro-5-(3-O-t-butyldimethylsilylpropoxy)benzaldehyde (1.02 g, 3 mmol) was dissolved 50 ml of anhydrous methylene chloride. 2 M Trimethylaluminum in

-229-

toluene (3 ml) was added dropwise within 10 min and kept the reaction mixture at room temperature. It was stirred further for 10 min and the mixture was poured into 10 ml ice cooled water. The emulsion was separated from water phase and dried over 100 g of sodium sulfate to remove the remaining water. The solvent was removed in vacuo and the mixture was applied to a silica gel column with gradient methanol in methylene chloride. 0.94 g (86%) of desired product was isolated.

$R_f = 0.375$  (hexane/ethyl acetate, 5/1).

UV (methanol), maximum: 306, 233, 206 nm; minimum: 255, 220 nm.

10  $^1\text{H}$  NMR ( $\text{DMSO}-d_6$ )  $\delta$  8.00 (d, 1H), 7.36 (s, 1H), 7.00 (d, 1H), 5.49 (b, OH), 5.31 (q, 1H), 4.19 (m, 2H), 3.77 (t, 2H), 1.95 (m, 2H), 1.37 (d, 3H), 0.86 (s, 9H), 0.04 (s, 6H).

$^{13}\text{C}$  NMR ( $\text{DMSO}-d_6$ )  $\delta$  162.6, 146.2, 139.6, 126.9, 112.9, 112.5, 64.8, 63.9, 58.7, 31.5, 25.6, 24.9, -3.4, -5.8.

15 D. 1-(2-Nitro-5-(3-hydroxypropoxy)phenyl)ethanol

1-(2-Nitro-5-(3-O-t-butyltrimethylsilylpropoxy)phenyl)ethanol (0.89 g, 2.5 mmol) was dissolved in 30 ml of THF and 0.5 mmol of  $n\text{Bu}_4\text{NF}$  was added under stirring. The mixture was stirred at room temperature for 5 h and the solvent was removed in vacuo. The remaining residue was applied to a silica gel column with gradient methanol in methylene chloride. 1-(2-Nitro-5-(3-hydroxypropoxy)phenyl)ethanol (0.6 g (99%)) was obtained.

$R_f = 0.17$  (dichloromethane/methanol, 95/5).

UV (methanol), maximum: 304, 232, 210 nm; minimum: 255, 219 nm.

25  $^1\text{H}$  NMR ( $\text{DMSO}-d_6$ )  $\delta$  8.00 (d, 1H), 7.33 (s, 1H), 7.00 (d, 1H), 5.50 (d, OH), 5.28 (t, OH), 4.59 (t, 1H), 4.17 (t, 2H), 3.57 (m, 2H), 1.89 (m, 2H), 1.36 (d, 2H).

$^{13}\text{C}$  NMR ( $\text{DMOS}-d_6$ )  $\delta$  162.8, 146.3, 139.7, 127.1, 113.1, 112.6, 65.5, 64.0, 57.0, 31.8, 25.0.

-230-

E. 1-(2-Nitro-5-(3-O-4,4'-dimethoxytritylpropoxy)phenyl)ethanol

1-(2-Nitro-5-(3-hydroxypropoxy)phenyl)ethanol (0.482 g, 2 mmol) was co-evaporated with anhydrous pyridine twice and dissolved in 20 ml anhydrous pyridine. The solution was cooled in ice-water bath and 750 mg (2.2 mmol) of DMTCl was added. The reaction mixture was stirred at room temperature overnight and 0.5 ml methanol was added to stop the reaction. The solvent was removed in vacuo and the residue was co-evaporated with toluene twice to remove trace of pyridine. The final residue was applied to a silica gel column with gradient methanol in methylene chloride containing drops of triethylamine to yield 0.96 g (89%) of the desired product 1-(2-nitro-5-(3-O-4,4'-dimethoxytritylpropoxy)phenyl)ethanol.

$R_f$  = 0.50 (dichloromethane/methanol, 99/1).

UV (methanol), maximum: 350 (shoulder), 305, 283, 276 (shoulder), 233, 208 nm; minimum: 290, 258, 220 nm.

$^1\text{H}$  NMR (DMSO- $d_6$ )  $\delta$  8.00 (d, 1H), 6.82-7.42 (ArH), 5.52 (d, OH), 5.32 (m, 1H), 4.23 (t, 2H), 3.71 (s, 6H), 3.17 (t, 2H), 2.00 (m, 2H), 1.37 (d, 3H).

$^{13}\text{C}$  NMR (DMOS- $d_6$ )  $\delta$  162.5, 157.9, 157.7, 146.1, 144.9, 140.1, 139.7, 135.7, 129.5, 128.8, 127.6, 127.5, 127.3, 126.9, 126.4, 113.0, 112.8, 112.6, 85.2, 85.3, 63.9, 59.0, 54.8, 28.9, 24.9.

F. 1-(2-Nitro-5-(3-O-4,4'-dimethoxytritylpropoxy)phenyl)-1-O-((2-cyanoethoxy)-diisopropylaminophosphino)ethane

1-(2-Nitro-5-(3-O-4,4'-dimethoxytritylpropoxy)phenyl)ethanol (400 mg, 0.74 mmol) was dried under high vacuum and was dissolved in 20 ml of anhydrous methylene chloride. To this solution, it was added 0.5 ml N,N-diisopropylethylamine and 0.3 ml (1.34 mmol) of 2-cyanoethyl-N,N-diisopropylchlorophosphoramidite. The reaction mixture was stirred at room temperature for 30 min and 0.5 ml of methanol was added to

-231-

stop the reaction. The mixture was washed with saturated sodium bicarbonate solution and was dried over sodium sulfate. The solvent was removed in vacuo and a quick silica gel column with 1% methanol in methylene chloride containing drops of triethylamine yield 510 mg (93%)

5 the desired phosphoramidite.

$R_f = 0.87$  (dichloromethane/methanol, 99/1).

#### EXAMPLE 35

1-(4-(3-O-4,4'-Dimethoxytritylpropoxy)-3-methoxy-6-nitrophenyl)-1-O-((2-cyanoethoxy)-diisopropylaminophosphino)ethane

10 A. 4-(3-Hydroxypropoxy)-3-methoxyacetophenone

3-Bromo-1-propanol (53 ml, 33 mmol) was refluxed in 100 ml of anhydrous acetonitrile with 4-hydroxy-3-methoxyacetophenone (5 g, 30 mmol),  $K_2CO_3$  (5 g), and KI (300 mg) overnight (15 h).

Methylenechloride (150 ml) was added to the reaction mixture after  
15 cooling to room temperature. The mixture was filtered and the solid residue was washed with methylene chloride. The combined organic solution was evaporated to dryness and redissolved in 100 ml methylene chloride. The resulted solution was washed with saturated NaCl solution and dried over sodium sulfate. 6.5 g (96.4%) of desired product was  
20 obtained after removal of the solvent in vacuo.

$R_f = 0.41$  (dichloromethane/methanol, 95/5).

UV (methanol), maximum: 304, 273, 227, 210 nm; minimum: 291, 244, 214 nm.

$^1H$  NMR (DMSO- $d_6$ )  $\delta$  7.64 (d, 1H), 7.46 (s, 1H), 7.04 (d, 1H), 4.58 (b, OH), 4.12 (t, 2H), 3.80 (s, 3H), 3.56 (t, 2H), 2.54 (s, 3H), 1.88 (m, 2H).  
25  $^{13}C$  NMR (DMSO- $d_6$ )  $\delta$  196.3, 152.5, 148.6, 129.7, 123.1, 111.5, 110.3, 65.4, 57.2, 55.5, 31.9, 26.3.

B. 4-(3-Acetoxypropoxy)-3-methoxyacetophenone

-232-

4-(3-Hydroxypropoxy)-3-methoxyacetophenone (3.5 g, 15.6 mmol) was dried and dissolved in 80 ml anhydrous acetonitrile. This mixture, 6 ml of triethylamine and 6 ml of acetic anhydride were added. After 4 h, 6 ml methanol was added and the solvent was removed in vacuo. The residue was dissolved in 100 ml dichloromethane and the solution was washed with dilute sodium bicarbonate solution, then water. The organic phase was dried over sodium sulfate and the solvent was removed. The solid residue was applied to a silica gel column with methylene chloride to yield 4.1g of 4-(3-acetoxypoxy)-3-methoxyacetophenone (98.6%).  
 $R_f = 0.22$  (dichloromethane/methanol, 99/1).  
UV (methanol), maximum: 303, 273, 227, 210 nm; minimum: 290, 243, 214 nm.

$^1\text{H}$  NMR (DMSO- $d_6$ )  $\delta$  7.62 (d, 1H), 7.45 (s, 1H), 7.08 (d, 1H), 4.12 (m, 4H), 3.82 (s, 3H), 2.54 (s, 3H), 2.04 (m, 2H), 2.00 (s, 3H).  
 $^{13}\text{C}$  NMR (DMSO- $d_6$ )  $\delta$  196.3, 170.4, 152.2, 148.6, 130.0, 123.0, 111.8, 110.4, 65.2, 60.8, 55.5, 27.9, 26.3, 20.7.

C. 4-(3-Acetoxypoxy)-3-methoxy-6-nitroacetophenone  
4-(3-Acetoxypoxy)-3-methoxyacetophenone (3.99 g, 15 mmol) was added portionwise to 15 ml of 70%  $\text{HNO}_3$  in water bath and keep the reaction temperature at the room temperature. The reaction mixture was stirred at room temperature for 30 min and 30 g of crushed ice was added. This mixture was extracted with 100 ml of dichloromethane and the organic phase was washed with saturated sodium bicarbonate solution. The solution was dried over sodium sulfate and the solvent was removed in vacuo. The crude mixture was applied to a silica gel column with gradient methanol in methylene chloride to yield 3.8 g (81.5%) of desired product 4-(3-acetoxypoxy)-3-methoxy-6-

-233-

nitroacetophenone and 0.38 g (8%) of ipso-substituted product 5-(3-acetoxypropoxy)-4-methoxy-1,2-dinitrobenzene.

Side ipso-substituted product 5-(3-acetoxypropoxy)-4-methoxy-1,2-dinitrobenzene:

5  $R_f = 0.47$  (dichloromethane/methanol, 99/1).

UV (methanol), maximum: 334, 330, 270, 240, 212 nm; minimum: 310, 282, 263, 223 nm.

$^1\text{H}$  NMR ( $\text{CDCl}_3$ )  $\delta$  7.36 (s, 1H), 7.34 (s, 1H), 4.28 (t, 2H), 4.18 (t, 2H), 4.02 (s, 3H), 2.20 (m, 2H), 2.08 (s, 3H).

10  $^{13}\text{C}$  NMR ( $\text{CDCl}_3$ )  $\delta$  170.9, 152.2, 151.1, 117.6, 111.2, 107.9, 107.1, 66.7, 60.6, 56.9, 28.2, 20.9.

Desired product 4-(3-acetoxypropoxy)-3-methoxy-6-nitroacetophenone:

$R_f = 0.29$  (dichloromethane/methanol, 99/1).

UV (methanol), maximum: 344, 300, 246, 213 nm; minimum: 320,

15 270, 227 nm.

$^1\text{H}$  NMR ( $\text{CDCl}_3$ )  $\delta$  7.62 (s, 1H), 6.74 (s, 1H), 4.28 (t, 2H), 4.20 (t, 2H), 3.96 (s, 3H), 2.48 (s, 3H), 2.20 (m, 2H), 2.08 (s, 3H).

$^{13}\text{C}$  NMR ( $\text{CDCl}_3$ )  $\delta$  200.0, 171.0, 154.3, 148.8, 138.3, 133.0, 108.8, 108.0, 66.1, 60.8, 56.6, 30.4, 28.2, 20.9.

20 D. 1-(4-(3-Hydroxypropoxy)-3-methoxy-6-nitrophenyl)ethanol

4-(3-Acetoxypropoxy)-3-methoxy-6-nitroacetophenone (3.73 g, 12

mmol) was added 150 ml ethanol and 6.5 g of  $\text{K}_2\text{CO}_3$ . The mixture was stirred at room temperature for 4h and TLC with 5% methanol in

dichloromethane indicated the completion of the reaction. To this same

25 reaction mixture, it was added 3.5 g of  $\text{NaBH}_4$  and the mixture was stirred at room temperature for 2h. Acetone (10 ml) was added to react with the remaining  $\text{NaBH}_4$ . The solvent was removed in vacuo and the residue was uptaken into 50 g of silica gel. The silica gel mixture was applied on the top of a silica gel column with 5% methanol in methylene

-234-

chloride to yield 3.15 g (97%) of desired product 1-(4-(3-hydroxypropoxy)-3-methoxy-6-nitrophenyl)ethanol.

Intermediate product 4-(3-hydroxypropoxy)-3-methoxy-6-nitroacetophenone after deprotection:

5  $R_f = 0.60$  (dichloromethane/methanol, 95/5).

Final product 1-(4-(3-hydroxypropoxy)-3-methoxy-6-nitrophenyl)ethanol:

$R_f = 0.50$  (dichloromethane/methanol, 95/5).

UV (methanol), maximum: 344, 300, 243, 219 nm; minimum: 317, 264, 233 nm.

10  $^1\text{H}$  NMR ( $\text{DMSO-d}_6$ )  $\delta$  7.54 (s, 1H), 7.36 (s, 1H), 5.47 (d, OH), 5.27 (m, 1H), 4.55 (t, OH), 4.05 (t, 2H), 3.90 (s, 3H), 3.55 (q, 2H), 1.88 (m, 2H), 1.37 (d, 3H).

$^{13}\text{C}$  NMR ( $\text{DMSO-d}_6$ )  $\delta$  153.4, 146.4, 138.8, 137.9, 109.0, 108.1, 68.5, 65.9, 57.2, 56.0, 31.9, 29.6.

15 E. 1-(4-(3-O-4,4'-Dimethoxytritylpropoxy)-3-methoxy-6-nitrophenyl)ethanol

1-(4-(3-Hydroxypropoxy)-3-methoxy-6-nitrophenyl)ethanol (0.325

g, 1.2 mmol) was co-evaporated with anhydrous pyridine twice and dissolved in 15 ml anhydrous pyridine. The solution was cooled in ice-

20 water bath and 450 mg (1.33 mmol) of DMTCI was added. The reaction mixture was stirred at room temperature overnight and 0.5 ml methanol was added to stop the reaction. The solvent was removed in vacuo and the residue was co-evaporated with toluene twice to remove trace of

pyridine. The final residue was applied to a silica gel column with  
25 gradient methanol in methylene chloride containing drops of triethylamine to yield 605 mg (88%) of desired product 1-(4-(3-O-4,4'-dimethoxytritylpropoxy)-3-methoxy-6-nitrophenyl)ethanol.

$R_f = 0.50$  (dichloromethane/methanol, 95/5).



-235-

UV (methanol), maximum: 354, 302, 282, 274, 233, 209 nm; minimum: 322, 292, 263, 222 nm.

<sup>1</sup>H NMR (DMSO-d<sub>6</sub>) δ 7.54 (s, 1H), 6.8-7.4 (ArH), 5.48 (d, OH), 5.27 (m, 1H), 4.16 (t, 2H), 3.85 (s, 3H), 3.72 (s, 6H), 3.15 (t, 2H), 1.98 (t, 2H), 1.37 (d, 3H).

<sup>13</sup>C NMR (DMSO-d<sub>6</sub>) δ 157.8, 153.3, 146.1, 144.9, 138.7, 137.8, 135.7, 129.4, 128.7, 127.5, 127.4, 126.3, 112.9, 112.6, 108.9, 108.2, 85.1, 65.7, 63.7, 59.2, 55.8, 54.8, 29.0, 25.0.

10 F. 1-(4-(3-O-4,4'-Dimethoxytritylpropoxy)-3-methoxy-6-nitrophenyl)-1-O-((2-cyanoethoxy)-diisopropylaminophosphino)ethane

1-(4-(3-O-4,4'-Dimethoxytritylpropoxy)-3-methoxy-6-nitrophenyl)ethanol (200 mg, 3.5 mmol) was dried under high vacuum and was dissolved in 15 ml of anhydrous methylene chloride. To this solution, it was added 0.5 ml N,N-diisopropylethylamine and 0.2 ml (0.89 mmol) of 2-cyanoethyl-N,N-diisopropylchlorophosphoramidite. The reaction mixture was stirred at room temperature for 30 min and 0.5 ml of methanol was added to stop the reaction. The mixture was washed with saturated sodium bicarbonate solution and was dried over sodium sulfate. The solvent was removed in vacuo and a quick silica gel column with 1% methanol in methylene chloride containing drops of triethylamine yield 247 mg (91.3%) the desired phosphoramidite 1-(4-(3-O-4,4'-dimethoxytritylpropoxy)-3-methoxy-6-nitrophenyl)-1-O-((2-cyanoethoxy)-diisopropylaminophosphino)ethane.

25 R<sub>f</sub> = 0.87 (dichloromethane/methanol, 99/1).

-236-

## EXAMPLE 36

## Oligonucleotide synthesis

The oligonucleotide conjugates containing photocleavable linker were prepared by solid phase nucleic acid synthesis (see: Sinha *et al.* Tetrahedron Lett. 1983, 24, 5843-5846; Sinha *et al.* Nucleic Acids Res. 1984, 12, 4539-4557; Beaucage *et al.* Tetrahedron 1993, 49, 6123-6194; and Matteucci *et al.* J. Am. Chem. Soc. 1981, 103, 3185-3191) under standard conditions. In addition a longer coupling time period was employed for the incorporation of photocleavable unit and the 5' terminal amino group. The coupling efficiency was detected by measuring the absorbance of released DMT cation and the results indicated a comparable coupling efficiency of phosphoramidite 1-(2-nitro-5-(3-O-4,4'-dimethoxytritylpropoxy)phenyl)-1-O-((2-cyanoethoxy)-diisopropylaminophosphino)ethane or 1-(4-(3-O-4,4'-dimethoxytritylpropoxy)-3-methoxy-6-nitrophenyl)-1-O-((2-cyanoethoxy)-diisopropylaminophosphino)ethane with those of common nucleoside phosphoramidites. Deprotection of the base protection and release of the conjugates from the solid support was carried out with concentrated ammonium at 55 °C overnight. Deprotection of the base protection of other conjugates was done by fast deprotection with AMA reagents. Purification of the MMT-on conjugates was done by HPLC (trityl-on) using 0.1 M triethylammonium acetate, pH 7.0 and a gradient of acetonitrile (5% to 25% in 20 minutes). The collected MMT or DMT protected conjugate was reduced in volume, detritylated with 80% aqueous acetic acid (40 min, 0 °C), desalted, stored at -20°C.

-237-

**EXAMPLE 37****Photolysis study**

In a typical case, 2 nmol of oligonucleotide conjugate containing photocleavable linker in 200  $\mu$ l distilled water was irradiated with a long wavelength UV lamp (Blak Ray XX-15 UV lamp, Ultraviolet products, San Gabriel, CA) at a distance of 10 cm (emission peak 365 nm, lamp intensity = 1.1 mW/cm<sup>2</sup> at a distance of 31 cm). The resulting mixture was analyzed by HPLC (trityl-off) using 0.1 M triethylammonium acetate, pH 7.0 and a gradient of acetonitrile. Analysis showed that the conjugate was cleaved from the linker within minutes upon UV irradiation.

**Equivalents**

Those skilled in the art will recognize, or be able to ascertain using no more than routine experimentation, numerous equivalents to the specific procedures described herein. Such equivalents are considered to be within the scope of this invention and are covered by the following claims.

-238-

## SEQUENCE LISTING

## (1) GENERAL INFORMATION

## (i) APPLICANT:

(A) NAME: SEQUENOM, INC.  
(B) STREET: 11555 Sorrento Valley Road  
(C) CITY: San Diego  
(D) STATE: California  
(E) COUNTRY: USA  
(F) POSTAL CODE (ZIP): 92121

## (i) INVENTOR/APPLICANT:

(A) NAME: Hubert Köster  
(B) STREET: 836 Via Mallorca Drive  
(C) CITY: La Jolla  
(D) STATE: California  
(E) COUNTRY: USA  
(F) POSTAL CODE (ZIP): 92037

## (i) INVENTOR/APPLICANT:

(A) NAME: Kai Tang  
(B) STREET: 8521 Summerdale Rd #241  
(C) CITY: San Diego  
(D) STATE: California  
(E) COUNTRY: USA  
(F) POSTAL CODE (ZIP): 92126

## (i) INVENTOR/APPLICANT:

(A) NAME: Cong-Jing Fu  
(B) STREET: 10615 Dabney Dr. #21  
(C) CITY: San Diego  
(D) STATE: California  
(E) COUNTRY: USA  
(F) POSTAL CODE (ZIP): 92126

## (i) INVENTOR/APPLICANT:

(A) NAME: Carsten W. Siegert  
(B) STREET: Geilstr. 42  
(C) CITY: 22303 Hamburg  
(D) STATE:  
(E) COUNTRY: Germany  
(F) POSTAL CODE (ZIP):

## (i) INVENTOR/APPLICANT:

(A) NAME: Daniel P. Little  
(B) STREET: 393 Glendale Lake Rd.  
(C) CITY: Patton  
(D) STATE: Pennsylvania  
(E) COUNTRY: USA  
(F) POSTAL CODE (ZIP): 18668

## (i) INVENTOR/APPLICANT:

(A) NAME: G. Scott Higgins  
(B) STREET: Haselweg 1  
(C) CITY: 22880 Weidel  
(D) STATE:  
(E) COUNTRY: Germany  
(F) POSTAL CODE (ZIP):

-239-

## (i) INVENTOR/APPLICANT:

(A) NAME: Andreas Braun  
(B) STREET: 13232 Benchley Road  
(C) CITY: San Diego  
(D) STATE: California  
(D) COUNTRY: USA  
(E) POSTAL CODE (ZIP): 92130

## (i) INVENTOR/APPLICANT:

(A) NAME: Brigitte Damhoffer-Demar  
(B) STREET: 3899 Haines St. #8-308  
(C) CITY: San Diego  
(D) STATE: California  
(D) COUNTRY: USA  
(E) POSTAL CODE (ZIP): 92109

## (i) INVENTOR/APPLICANT:

(A) NAME: Christian Jurinke  
(B) STREET: Grope Hall 68  
(C) CITY: 22115 Hamburg  
(D) STATE:  
(D) COUNTRY: Germany  
(E) POSTAL CODE (ZIP):

## (i) INVENTOR/APPLICANT:

(A) NAME: Dirk Van den Boom  
(B) STREET: Forsthausstr. 8  
(C) CITY: 633303 Preiech  
(D) STATE:  
(D) COUNTRY: Germany  
(E) POSTAL CODE (ZIP):

## (i) INVENTOR/APPLICANT:

(A) NAME: Goubing Xiang  
(B) STREET: 11381 Zapata Ave.  
(C) CITY: San Diego  
(D) STATE: California  
(D) COUNTRY: USA  
(E) POSTAL CODE (ZIP): 92126

(ii) TITLE OF THE INVENTION: DNA DIAGNOSTICS BASED ON MASS SPECTROMETRY

(iii) NUMBER OF SEQUENCES: 320

## (iv) CORRESPONDENCE ADDRESS:

(A) ADDRESSEE: Brown, Martin, Haller & McClain  
(B) STREET: 1660 Union Street  
(C) CITY: San Diego  
(D) STATE: CA  
(E) COUNTRY: USA  
(F) ZIP: 92101-2926

## (v) COMPUTER READABLE FORM:

(A) MEDIUM TYPE: Diskette  
(B) COMPUTER: IBM Compatible  
(C) OPERATING SYSTEM: DOS  
(D) SOFTWARE: FastSEQ Version 1.5

## (vi) CURRENT APPLICATION DATA:

(A) APPLICATION NUMBER:

-240-

- (B) FILING DATE: 06-NOV-1997
- (C) CLASSIFICATION:
- (vii) PRIOR APPLICATION DATA:
  - (A) APPLICATION NUMBER:
  - (B) FILING DATE: 10/08/97
- (vii) PRIOR APPLICATION DATA:
  - (A) APPLICATION NUMBER: 08/933,792
  - (B) FILING DATE: 09/19/97
- (vii) PRIOR APPLICATION DATA:
  - (A) APPLICATION NUMBER: 08/787,639
  - (B) FILING DATE: 01/23/97
- (vii) PRIOR APPLICATION DATA:
  - (A) APPLICATION NUMBER: 08/786,988
  - (B) FILING DATE: 01/23/97
- (vii) PRIOR APPLICATION DATA:
  - (A) APPLICATION NUMBER: 08/746,055
  - (B) FILING DATE: 11/06/96
- (vii) PRIOR APPLICATION DATA:
  - (A) APPLICATION NUMBER: 08/746,036
  - (B) FILING DATE: 11/06/96
- (vii) PRIOR APPLICATION DATA:
  - (A) APPLICATION NUMBER: 08/744,590
  - (B) FILING DATE: 11/06/96
- (vii) PRIOR APPLICATION DATA:
  - (A) APPLICATION NUMBER: 08/744,481
  - (B) FILING DATE: 11/06/96
- (viii) ATTORNEY/AGENT INFORMATION:
  - (A) NAME: Seidman, Stephanie L
  - (B) REGISTRATION NUMBER: 33,779
  - (C) REFERENCE/DOCKET NUMBER: 7352-2004PC
- (ix) TELECOMMUNICATION INFORMATION:
  - (A) TELEPHONE: 619-238-0999
  - (B) TELEFAX: 619-238-0062
  - (C) TELEX:
- (2) INFORMATION FOR SEQ ID NO:1:
  - (i) SEQUENCE CHARACTERISTICS:
    - (A) LENGTH: 20 base pairs
    - (B) TYPE: nucleic acid
    - (C) STRANDEDNESS: single
    - (D) TOPOLOGY: unknown
  - (ii) MOLECULE TYPE: cDNA
  - (iii) HYPOTHETICAL: NO
  - (iv) ANTISENSE: NO
  - (v) FRAGMENT TYPE:
  - (vi) ORIGINAL SOURCE:
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

-241-

GCAAGTGAAT CCTGAGCGTG

20

## (2) INFORMATION FOR SEQ ID NO:2:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 19 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

GTGTGAAGGG TTCATATGC

19

## (2) INFORMATION FOR SEQ ID NO:3:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 28 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

ATCTATATTC ATCATAGGAA ACACCACA

28

## (2) INFORMATION FOR SEQ ID NO:4:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 30 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

GTATCTATAT TCATCATAGG AAACACCATT

30

## (2) INFORMATION FOR SEQ ID NO:5:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 30 base pairs
- (B) TYPE: nucleic acid

-242-

- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

GCTTTGGGGC ATGGACATTG ACCCGTATAA

30

- (2) INFORMATION FOR SEQ ID NO:6:

- (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 30 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

CTGACTACTA ATCCCTGGA TGCTGGGTCT

30

- (2) INFORMATION FOR SEQ ID NO:7:

- (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

TTGCCTGAGT GCAGTATGGT

20

- (2) INFORMATION FOR SEQ ID NO:8:

- (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO



-243-

- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

AGCTCTATAT CGGGAAGCCT

20

(2) INFORMATION FOR SEQ ID NO:9:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 24 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

TTGTGCCACG CGGTTGGGAA TGTA

24

(2) INFORMATION FOR SEQ ID NO:10:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 26 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

AGCAACGACT GTTTGCCCGC CAGTTG

26

(2) INFORMATION FOR SEQ ID NO:11:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 25 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

-244-

TACATTCCCA ACCGCGTGGC ACAAC

25

## (2) INFORMATION FOR SEQ ID NO:12:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 25 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

AACTGGCGGG CAAACAGTCG TTGCT

25

## (2) INFORMATION FOR SEQ ID NO:13:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

GCAAGTGAAT CCTGAGCGTG

20

## (2) INFORMATION FOR SEQ ID NO:14:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 14 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

GTGTGAAGGG CGTG

14

## (2) INFORMATION FOR SEQ ID NO:15:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 24 base pairs

-245-

- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

CTATATTCAT CATAGGAAAC ACCA

24

(2) INFORMATION FOR SEQ ID NO:16:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 18 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

GTCACCCTCG ACCTCCAG

18

(2) INFORMATION FOR SEQ ID NO:17:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 19 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

TTGTAAAACG ACGCCAGT

19

(2) INFORMATION FOR SEQ ID NO:18:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 18 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO

-246-

- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

CTTCCACCGC GATGTTGA

18

(2) INFORMATION FOR SEQ ID NO:19:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 17 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

CAGGAAACAG CTATGAC

17

(2) INFORMATION FOR SEQ ID NO:20:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 17 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

GTAAAACGAC GGCCAGT

17

(2) INFORMATION FOR SEQ ID NO:21:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 19 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

GTCACCTCG ACCTGCAGC

19

-247-

## (2) INFORMATION FOR SEQ ID NO:22:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

--(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

GTTGTAAAC GAGGCCAGT

20

## (2) INFORMATION FOR SEQ ID NO:23:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 39 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

TCTGGCCTGG TGCAGGCCT ATTGTAGTTG TGACGTACA

39

## (2) INFORMATION FOR SEQ ID NO:24:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 14 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

TGTACGTCAC AACT

14

## (2) INFORMATION FOR SEQ ID NO:25:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 78 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

-248-

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:

AAGATCTGAC CAGGGATTTCG GTTAGCGTGA CTGCTGCTGC TGCTGCTGCT GCTGGATGAT 60  
CCGACGCATC AGATCTGG

(2) INFORMATION FOR SEQ ID NO:26:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 18 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:

CTGATGCGTC GGATCATC

18

(2) INFORMATION FOR SEQ ID NO:27:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 23 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:

GATGATCCGA CGCATCACAG CTC

23

(2) INFORMATION FOR SEQ ID NO:28:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 33 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

-249-

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:28:

TCGGTTCCAA GAGCTGTGAT GCGTCGGATC ATC

33

(2) INFORMATION FOR SEQ ID NO:29:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 23 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:29:

GATGATCCGA CGCATCACAG CTC

23

(2) INFORMATION FOR SEQ ID NO:30:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 18 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:30:

GTGATGCGTC GGATCATC

18

(2) INFORMATION FOR SEQ ID NO:31:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 15 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:31:

TCGGTTCCAA GAGCT

15

(2) INFORMATION FOR SEQ ID NO:31:

(i) SEQUENCE CHARACTERISTICS:

-250-

- (A) LENGTH: 15 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:31:

TCGGTTCCAA GAGCT

15

(2) INFORMATION FOR SEQ ID NO:32:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 21 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:32:

CATTGCTTC TGACACAACT G

21

(2) INFORMATION FOR SEQ ID NO:33:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 19 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:33:

CTTCTCTGTC TCCACATGC

19

(2) INFORMATION FOR SEQ ID NO:34:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 12 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO



-251-

- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:34:

TGCACCTGAC TC

12

(2) INFORMATION FOR SEQ ID NO:35:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 20 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:35:

TGCTTACTTA ACCCAGTGTG

20

(2) INFORMATION FOR SEQ ID NO:36:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 20 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:36:

CACACTATGT AATACTATGC

20

(2) INFORMATION FOR SEQ ID NO:37:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 22 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:37:

-252-

GAAAATATCT GACAAACTCA TC

22

(2) INFORMATION FOR SEQ ID NO:38:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 21 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:38:

CATGGACACC AAATTAAGTT C

21

(2) INFORMATION FOR SEQ ID NO:39:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 14 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:39:

TGAGACTCTG TCTC

14

(2) INFORMATION FOR SEQ ID NO:40:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 15 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:40:

TTCCCCAAAT CCCTG

15

(2) INFORMATION FOR SEQ ID NO:41:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 19 base pairs

-253-

- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:41:

GGCACGGCTG TCCAAGGAG

19

(2) INFORMATION FOR SEQ ID NO:42:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 20 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:42:

AGGCCGCGCT CGGCGCCCTC

20

(2) INFORMATION FOR SEQ ID NO:43:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 18 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:43:

GCGGACATGG AGGACGTG

18

(2) INFORMATION FOR SEQ ID NO:44:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 21 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO

-254-

- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:44:

GATGCCGATG ACCTGCAGAA G

21

(2) INFORMATION FOR SEQ ID NO:45:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 24 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:45:

CCCTTACCCT TACCCTTACC CTAA

24

(2) INFORMATION FOR SEQ ID NO:46:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 18 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:46:

AATCCGTGCA GCAGAGTT

18

(2) INFORMATION FOR SEQ ID NO:47:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:47:

TGTCAGAGCT GGACAAGTGT

20

-255-

## (2) INFORMATION FOR SEQ ID NO:48:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:48:

GATATTGTCT TCCCGGTAGC

20

## (2) INFORMATION FOR SEQ ID NO:49:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:49:

CTCGGACCAG GTGTACCGCC

20

## (2) INFORMATION FOR SEQ ID NO:50:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 21 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:50:

CCTGTACTGG AAGGCGATCT C

21

## (2) INFORMATION FOR SEQ ID NO:51:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

-256-

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:51:

CATGAGGCAG AGCATAACGCA

20

(2) INFORMATION FOR SEQ ID NO:52:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 20 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:52:

GACAGCAGCA CCGAGACGAT

20

(2) INFORMATION FOR SEQ ID NO:53:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 20 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:53:

CGGCTGCGAT CACCGTGCGG

20

(2) INFORMATION FOR SEQ ID NO:54:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 19 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:54:

-257-

GATCCACTGT GCGACGAGC

19

(2) INFORMATION FOR SEQ ID NO:55:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:55:

GCGGCTGCGA TCACCGTGC

19

(2) INFORMATION FOR SEQ ID NO:56:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:56:

TGCACCTGAC TC

12

(2) INFORMATION FOR SEQ ID NO:57:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 12 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:57:

CTGTGGTCGT GC

12

(2) INFORMATION FOR SEQ ID NO:58:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 39 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single

-258-

(D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:58:

GAGTCAGGTG CGCCATGCCT CAAACAGACA CCATGGCGC

39

(2) INFORMATION FOR SEQ ID NO:59:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 21 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:59:

TCTCTGTCTC CACATGCCCA G

21

(2) INFORMATION FOR SEQ ID NO:60:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:60:

ACCTAGCGTT CAGTTCGACT GAGATAATAC GACTCACTAT AGCAGCTCTC ATTTCCATA 60  
C 61

(2) INFORMATION FOR SEQ ID NO:61:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 21 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:



-259-

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:61:

AACTAAGCCA TGTGCACAAC A

21

(2) INFORMATION FOR SEQ ID NO:62:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 25 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: RNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:62:

UCCGGUCUGA UGAGUCCGUG AGGAC

25

(2) INFORMATION FOR SEQ ID NO:63:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: RNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:63:

GUCACUACAG GUGAGCUCCA

20

(2) INFORMATION FOR SEQ ID NO:64:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: RNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:64:

CCAUGCGAGA GUAAGUAGUA

20

(2) INFORMATION FOR SEQ ID NO:65:

-260-

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

## (ii) MOLECULE TYPE: RNA

## (iii) HYPOTHETICAL: NO

## (iv) ANTISENSE: NO

## (v) FRAGMENT TYPE:

## (vi) ORIGINAL SOURCE:

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:65:

AGGCCUGCGG CAAGACGGAA AGACCAUGGU CCCNAUCUG CCGCAGGAUC

50

## (2) INFORMATION FOR SEQ ID NO:65:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

## (ii) MOLECULE TYPE: cDNA

## (iii) HYPOTHETICAL: NO

## (iv) ANTISENSE: NO

## (v) FRAGMENT TYPE:

## (vi) ORIGINAL SOURCE:

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:66:

CATTGCTTC TGACACAACT

20

## (2) INFORMATION FOR SEQ ID NO:67:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 21 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

## (ii) MOLECULE TYPE: cDNA

## (iii) HYPOTHETICAL: NO

## (iv) ANTISENSE: NO

## (v) FRAGMENT TYPE:

## (vi) ORIGINAL SOURCE:

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:67:

TCTCTGTCTC CACATGCCCA G

21

## (2) INFORMATION FOR SEQ ID NO:68:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 24 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

## (ii) MOLECULE TYPE: cDNA

-261-

- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:68:

GTCGTCCCAT GGTGCACCTG ACTC

24

--(2) INFORMATION FOR SEQ ID NO:69:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 22 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:69:

CGCTGTGGTG AGGCCCTGGG CA

22

(2) INFORMATION FOR SEQ ID NO:70:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 24 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:70:

GACGACGACT GCTACCTGAC TCCA

24

(2) INFORMATION FOR SEQ ID NO:71:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 24 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:71:

-262-

ACAGCGGACT GCTACCTGAC TCCA

24

(2) INFORMATION FOR SEQ ID NO:72:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 18 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:72:

TGGAGTCAGG TAGCAGTC

18

(2) INFORMATION FOR SEQ ID NO:73:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 19 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:73:

CAGCTCTCAT TTCCATAC

19

(2) INFORMATION FOR SEQ ID NO:74:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 18 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:74:

AGCCCCAAGA TGACTATC

18

(2) INFORMATION FOR SEQ ID NO:75:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 21 base pairs
- (B) TYPE: nucleic acid

-263-

- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:75:

CGAGGAGCTC AAGGCCAGAA T

21

(2) INFORMATION FOR SEQ ID NO:76:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 18 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:76:

CAGGGGCAGC TCAGCTCTC

19

(2) INFORMATION FOR SEQ ID NO:77:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 18 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:77:

GGCACGGCTG TCCAAGGA

18

(2) INFORMATION FOR SEQ ID NO:78:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 18 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:

-264-

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:78

AGGCCGCGCT CGGCGCCCTC

20

(2) INFORMATION FOR SEQ ID NO:79:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 18 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:79:

CTTACTTGAA TTCCAAGAGC

20

(2) INFORMATION FOR SEQ ID NO:80:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 22 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:80:

GGGCTGACTT GCATGGACCG GA

22

(2) INFORMATION FOR SEQ ID NO:81

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 12 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:81:

AGCCAGGACA AG

12

-265-

## (2) INFORMATION FOR SEQ ID NO:82:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 15 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:82:

ACAGCAGGAA CAGCA

15

## (2) INFORMATION FOR SEQ ID NO:83:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 18 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:83:

GCGGACATGG AGGACGTG

18

## (2) INFORMATION FOR SEQ ID NO:84:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 21 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:84:

GATGCCGATG ACCTGCAGAA G

21

## (2) INFORMATION FOR SEQ ID NO:85:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 24 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

-266-

(ii) MOLECULE TYPE: cDNA  
(iii) HYPOTHETICAL: NO  
(iv) ANTISENSE: NO  
(v) FRAGMENT TYPE:  
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:85:

GTGCCCTGCA GCTTCACTGA AGAC

24

(2) INFORMATION FOR SEQ ID NO:86:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 12 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA  
(iii) HYPOTHETICAL: NO  
(iv) ANTISENSE: NO  
(v) FRAGMENT TYPE:  
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:86:

AGCCAGGACA AG

12

(2) INFORMATION FOR SEQ ID NO:87:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 14 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA  
(iii) HYPOTHETICAL: NO  
(iv) ANTISENSE: NO  
(v) FRAGMENT TYPE:  
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:87:

AGCCAGGACA AGTC

14

(2) INFORMATION FOR SEQ ID NO:88:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 13 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA  
(iii) HYPOTHETICAL: NO  
(iv) ANTISENSE: NO  
(v) FRAGMENT TYPE:  
(vi) ORIGINAL SOURCE:



-267-

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:88:

AGCCAGGACA AGA

13

(2) INFORMATION FOR SEQ ID NO:89:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 15 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:89:

ACAGCACCAA CAGCA

15

(2) INFORMATION FOR SEQ ID NO:90:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 17 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:90:

ACAGCAGGAA CAGCATC

17

(2) INFORMATION FOR SEQ ID NO:91:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 16 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:91:

ACAGCAGGAA CAGCAG

16

-268-

## (2) INFORMATION FOR SEQ ID NO:92:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 18 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:92:

GCGGACATGG AGGACGTG

18

## (2) INFORMATION FOR SEQ ID NO:93:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:93:

GCGGACATGG AGGACGTGGC

20

## (2) INFORMATION FOR SEQ ID NO:94:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 19 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:94:

GCGGACATGG AGGACGTGC

19

## (2) INFORMATION FOR SEQ ID NO:95:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 21 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single

-269-

(D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:95:

GATGCCGATG ACCTGCAGAA G

21

(2) INFORMATION FOR SEQ ID NO:96:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 22 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:96:

GATGCCGATG ACCTGCAGAA GC

22

(2) INFORMATION FOR SEQ ID NO:97:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 23 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:97:

GATGCCGATG ACCTGCAGAA GTG

23

(2) INFORMATION FOR SEQ ID NO:98:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 24 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:

-270-

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:98:

GTGCCCTGCA GCTTCACTGA AGAC

24

(2) INFORMATION FOR SEQ ID NO:99:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 26 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:99:

GTGCCCTGCA GCTTCACTGA AGACTG

26

(2) INFORMATION FOR SEQ ID NO:100:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 25 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:100:

GTGCCCTGCA GCTTCACTGA AGACC

25

(2) INFORMATION FOR SEQ ID NO:101:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 19 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:101:

TATCTGTCA CTTGTGCCC

19

-271-

## (2) INFORMATION FOR SEQ ID NO:102:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 19 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:102:

CAGAGGCCTG GGGACCTG

19

## (2) INFORMATION FOR SEQ ID NO:103:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 18 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:103:

ACGACAGGGC TGGTTGCC

18

## (2) INFORMATION FOR SEQ ID NO:104:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 19 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:104:

ACTGACAACC ACCCTTAAC

19

## (2) INFORMATION FOR SEQ ID NO:105:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 18 base pairs
- (B) TYPE: nucleic acid

-272-

- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:105:

CTGCTTGCCA CAGGTCTC

18

(2) INFORMATION FOR SEQ ID NO:106:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 18 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:106:

CACAGCAGGC CAGTGTCC

18

(2) INFORMATION FOR SEQ ID NO:107:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 19 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:107:

GGACCTGATT TCCTTACTG

19

(2) INFORMATION FOR SEQ ID NO:108:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 19 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA

-273-

- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:108:

TGAATCTGAG GCATAACTG

19

(2) INFORMATION FOR SEQ ID NO:109:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:109:

TTGCGTACAC ACTGGCCGTC GTTTTACAAC GTCGTGACTG GGAAAACCT

50

(2) INFORMATION FOR SEQ ID NO:110:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 27 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:110:

GTAAAACGAC GGCCAGTGTG TACGCAA

27

(2) INFORMATION FOR SEQ ID NO:111:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 27 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

-274-

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:111:

TACTGGAAGG CGATCTCAGC AATCAGC

27

(2) INFORMATION FOR SEQ ID NO:112:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 19 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:112:

GGCACGGCTG TCCAAGGAG

19

(189) INFORMATION FOR SEQ ID NO:113:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:113:

AGGCCGCGCT CGGCGCCCTC

20

(2) INFORMATION FOR SEQ ID NO:114:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: RNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:114:

GUCACUACAG GUGAGCUCCA

20

(2) INFORMATION FOR SEQ ID NO:115:



-275-

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 21 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:115:

GAATTCGAGC TCGGTACCCG G

21

## (2) INFORMATION FOR SEQ ID NO:116:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 21 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:116:

CCGGGTACCG AGCTCGAATT C

21

## (2) INFORMATION FOR SEQ ID NO:117:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 23 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: unknown
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:117:

CCTCTTGGA ACTGTGTAGT ATT

23

## (2) INFORMATION FOR SEQ ID NO:118:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 112 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

-276-

- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:118:

AGGCTGTCTC TCTCCCTCTC TCATACACAC ACACACACAC ACACACACAC ACACACACAC 60  
ACACACACAC TCACACTCAC CCACANNNA AATACTACACA GTTCCAAGA GG 112

(2) INFORMATION FOR SEQ ID NO:119:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 49 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:119:

TAATACGACT CACTATAGGG CGAAGGCTGT CTCTCTCCCT CTCTCATAC 49

(2) INFORMATION FOR SEQ ID NO:120:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 135 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:120:

TAATACGACT CACTATAGGG CGAAGGCTGT CTCTCTCCCT CTCTCATACA CACACACACA 60  
CACACACACA CACACACACA CACTCACACT CACCCACANN NAAATACTAC 120  
ACAGTTCCA AGAGG 135

(2) INFORMATION FOR SEQ ID NO:121:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 12 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

-277-

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:121:

AATACTACAC AG

12

(2) INFORMATION FOR SEQ ID NO:122:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 24 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: unknown
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:122:

CTGATGCGTC GGATCATCTT TTTT

24

(2) INFORMATION FOR SEQ ID NO:123:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 23 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:123:

GATGATCCGA CGCATCAGAA TGT

23

(2) INFORMATION FOR SEQ ID NO:124:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 29 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: unknown
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:124:

GATCTAGCTG GGCCGAGCTA GGCCGTTGA

29

(2) INFORMATION FOR SEQ ID NO:125:

(i) SEQUENCE CHARACTERISTICS:

-278-

- (A) LENGTH: 27 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:125:

CTGATGCGTC GGATCATCTT TTTTTT

27

(2) INFORMATION FOR SEQ ID NO:126:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 12 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:126:

GATGATCCGA CG

12

(2) INFORMATION FOR SEQ ID NO:127:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 15 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:127:

GATGATCCGA CGCAT

15

(2) INFORMATION FOR SEQ ID NO:128:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 12 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO

-279-

(iv) ANTISENSE: NO  
 (v) FRAGMENT TYPE:  
 (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:128:

AAAAAAGATG AT

12

--(2) INFORMATION FOR SEQ ID NO:129:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 12 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA  
 (iii) HYPOTHETICAL: NO  
 (iv) ANTISENSE: NO  
 (v) FRAGMENT TYPE:  
 (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:129:

GATCCGACGC AT

12

(2) INFORMATION FOR SEQ ID NO:130:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 253 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA  
 (iii) HYPOTHETICAL: NO  
 (iv) ANTISENSE: NO  
 (v) FRAGMENT TYPE:  
 (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:130:

GGCACGGCTG	TCCAAGGAGC	TGCAGGCGGC	GCAGGCCCGG	CTGGGCGCGG	ACATGGAGGA	60
CGTGTGCGGC	CGCCTGGTGC	AGTACCGCGG	CGAGGTGCAG	GCCATGCTCG	GCCAGAGCAC	120
CGAGGAGCTG	CGGGTGCGCC	TCGCCTCCCA	CCTGCGCAAG	CTGCGTAAGC	GGCTCCTCCG	180
CGATGCCGAT	GACCTGCAGA	AGTGCCTGGC	AGTGTACCAG	GCCGGGGCCC	GCGAGGGCGC	240
CGAGCGCGGC	CTC					253

(2) INFORMATION FOR SEQ ID NO:131:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 58 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA  
 (iii) HYPOTHETICAL: NO  
 (iv) ANTISENSE: NO  
 (v) FRAGMENT TYPE:  
 (vi) ORIGINAL SOURCE:

-280-

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:131:

GAATTACATT CCCAACCGCG TGGCACAACA ACTGGCGGGC AAACAGTCGT TGCTGATT 58

(2) INFORMATION FOR SEQ ID NO:132:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 57 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:132:

ACCATTAAAG AAAATATCAT CTTTGGTGTT TCCTATGATG AATATAGAAG CGTCATC 57

(2) INFORMATION FOR SEQ ID NO:133:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 29 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:133:

CTATATTCAT CATAGGAAAC ACCAAAGAT 29

(2) INFORMATION FOR SEQ ID NO:134:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 26 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:134:

CTATATTCAT CATAGGAAAC ACCAAT 26

(2) INFORMATION FOR SEQ ID NO:135:

(i) SEQUENCE CHARACTERISTICS:

-281-

- (A) LENGTH: 29 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:135:

CTATATTCAT CATAGGAAAC ACCAAAGAT

29

(2) INFORMATION FOR SEQ ID NO:136:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 38 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:136:

CTATATTCAT CATAGGAAAC ACCAAAGATG ATATTTTC

38

(2) INFORMATION FOR SEQ ID NO:137:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 35 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:137:

CTATATTCAT CATAGGAAAC ACCAATG ATATTTTC

35

(2) INFORMATION FOR SEQ ID NO:138:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 35 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO

-282-

- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:138:

CTATATTCAT CATAGGAAAC ACCAAAGATA TTTTC

35

(2) INFORMATION FOR SEQ ID NO:139:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 31 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:139:

CTATATTCAT CATAGGAAAC ACCAAAGATG C

31

(2) INFORMATION FOR SEQ ID NO:140:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 200 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:140:

CTTCCACCGC GATGTTGATG ATTATGTGTC TGAATTGAT GGGGGCAGGC GGCCCCCGTC  
TGTTTGTCGC GGGTCTGGTG TTGATGGTGG TTCCTGCCT TGTCACCCTC GACCTGCAGC  
CCAAGCTTGG GATCCACCAC CATCACCATC ACTAATAATG CATGGGCTGC AGCCAATTGG  
CACTGGCCGT CGTTTACAA

(2) INFORMATION FOR SEQ ID NO:141:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 99 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:



-283-

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:141:

GTCACCCTCG ACCTGCAGCC CAAGCTTGGG ATCCACCACC ATCACCATCA CTAATAATGC  
ATGGGCTGCA GCCAATTGGC ACTGGCCGTC GTTTTACAA

(2) INFORMATION FOR SEQ ID NO:142:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 15 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:142:

TGTACGTCAC AACTA

15

(2) INFORMATION FOR SEQ ID NO:143:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 16 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:143:

TGTACGTCAC AACTAC

16

(2) INFORMATION FOR SEQ ID NO:144:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 17 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:144:

TGTACGTCAC AACTACA

17

(2) INFORMATION FOR SEQ ID NO:145:

(i) SEQUENCE CHARACTERISTICS:

-284-

- (A) LENGTH: 18 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:145:

TGTACGTCAC AACTACAA

18

(2) INFORMATION FOR SEQ ID NO:146:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 19 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:146:

TGTACGTCAC AACTACAAT

19

(2) INFORMATION FOR SEQ ID NO:147:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:147:

TGTACGTCAC AACTACAATA

20

(2) INFORMATION FOR SEQ ID NO:148:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 21 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO

-285-

- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:148:

TGTACGTCAC AACTACAATA G

21

(2) INFORMATION FOR SEQ ID NO:149:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 22 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:149:

TGTACGTCAC AACTACAATA GG

22

(2) INFORMATION FOR SEQ ID NO:150:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 23 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:150:

TGTACGTCAC AACTACAATA GGC

23

(2) INFORMATION FOR SEQ ID NO:151:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 24 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:151:

TGTACGTCAC AACTACAATA GGCC

24

-286-

## (2) INFORMATION FOR SEQ ID NO:152:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 25 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:152:

TGTACGTCAC AACTACAATA GGCCC

25

## (2) INFORMATION FOR SEQ ID NO:153:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 26 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:153:

TGTACGTCAC AACTACAATA GGCCCT

26

## (2) INFORMATION FOR SEQ ID NO:154:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 27 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:154:

TGTACGTCAC AACTACAATA GGCCCTG

27

## (2) INFORMATION FOR SEQ ID NO:155:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 28 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

-287-

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:155:

TGTACGTCAC AACTACAATA GGCCCTGC

28

(2) INFORMATION FOR SEQ ID NO:156:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 29 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:156:

TGTACGTCAC AACTACAATA GGCCCTGCA

29

(2) INFORMATION FOR SEQ ID NO:157:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 30 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:157:

TGTACGTCAC AACTACAATA GGCCCTGCAC

30

(2) INFORMATION FOR SEQ ID NO:158:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 31 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:158:

-288-

TGTACGTCAC AACTACAATA GGCCCTGCAC C

31

## (2) INFORMATION FOR SEQ ID NO:159:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 32 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:159:

TGTACGTCAC AACTACAATA GGCCCTGCAC CA

32

## (2) INFORMATION FOR SEQ ID NO:160:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 33 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:160:

TGTACGTCAC AACTACAATA GGCCCTGCAC CAG

33

## (2) INFORMATION FOR SEQ ID NO:161:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 34 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:161:

TGTACGTCAC AACTACAATA GGCCCTGCAC CAGG

34

## (2) INFORMATION FOR SEQ ID NO:162:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 35 base pairs
- (B) TYPE: nucleic acid

-289-

- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:162:

TGTACGTCAC AACTACAATA GGCCCTGCAC CAGGC

35

(2) INFORMATION FOR SEQ ID NO:163:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 36 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:163:

TGTACGTCAC AACTACAATA GGCCCTGCAC CAGGCC

36

(2) INFORMATION FOR SEQ ID NO:164:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 37 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:164:

TGTACGTCAC AACTACAATA GGCCCTGCAC CAGGCCA

37

(2) INFORMATION FOR SEQ ID NO:165:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 38 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:

-290-

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:165:

TGTACGTCAC AACTACAATA GGCCCTGCAC CAGGCCAG

38

(2) INFORMATION FOR SEQ ID NO:166:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 39 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:166:

TGTACGTCAC AACTACAATA GGCCCTGCAC CAGGCCAGA

39

(2) INFORMATION FOR SEQ ID NO:167:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 19 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:167:

CTGATGCGTC GGATCATCC

19

(2) INFORMATION FOR SEQ ID NO:168:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:168:

CTGATGCGTC GGATCATCCA

20

(2) INFORMATION FOR SEQ ID NO:169:



-291-

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 21 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:169:

CTGATGCGTC GGATCATCCA G

21

## (2) INFORMATION FOR SEQ ID NO:170:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 22 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:170:

CTGATGCGTC GGATCATCCA GC

22

## (2) INFORMATION FOR SEQ ID NO:171:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 23 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:171:

CTGATGCGTC GGATCATCCA GCA

23

## (2) INFORMATION FOR SEQ ID NO:172:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 24 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

-292-

- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:172:

CTGATGCGTC GGATCATCCA GCAG

24

(2) INFORMATION FOR SEQ ID NO:173:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 25 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:173:

CTGATGCGTC GGATCATCCA GCAGC

25

(2) INFORMATION FOR SEQ ID NO:174:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 26 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:174:

CTGATGCGTC GGATCATCCA GCAGCA

26

(2) INFORMATION FOR SEQ ID NO:175:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 27 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:175:

-293-

CTGATGCGTC GGATCATCCA GCAGCAG

27

(2) INFORMATION FOR SEQ ID NO:176:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 28 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:176:

CTGATGCGTC GGATCATCCA GCAGCAGC

28

(2) INFORMATION FOR SEQ ID NO:177:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 29 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:177:

CTGATGCGTC GGATCATCCA GCAGCAGCA

29

(2) INFORMATION FOR SEQ ID NO:178:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 30 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:178:

CTGATGCGTC GGATCATCCA GCAGCAGCAG

30

(2) INFORMATION FOR SEQ ID NO:179:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 31 base pairs
- (B) TYPE: nucleic acid

-294-

- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:179:

CTGATGCGTC GGATCATCCA GCAGCAGCAG C

31

(2) INFORMATION FOR SEQ ID NO:180:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 32 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:180:

CTGATGCGTC GGATCATCCA GCAGCAGCAG CA

32

(2) INFORMATION FOR SEQ ID NO:181:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 33 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:181:

CTGATGCGTC GGATCATCCA GCAGCAGCAG CAG

33

(2) INFORMATION FOR SEQ ID NO:182:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 34 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:

-295-

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:182:

CTGATGCGTC GGATCATCCA GCAGCAGCAG CAGC

34

(2) INFORMATION FOR SEQ ID NO:183:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 35 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:183:

CTGATGCGTC GGATCATCCA GCAGCAGCAG CAGCA

35

(2) INFORMATION FOR SEQ ID NO:184:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 36 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:184:

CTGATGCGTC GGATCATCCA GCAGCAGCAG CAGCAG

36

(2) INFORMATION FOR SEQ ID NO:185:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 37 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:185:

CTGATGCGTC GGATCATCCA GCAGCAGCAG CAGCAGC

37

(2) INFORMATION FOR SEQ ID NO:186:

-296-

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 38 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:186:

CTGATGCGTC GGATCATCCA GCAGCAGCAG CAGCAGCA

38

## (2) INFORMATION FOR SEQ ID NO:187:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 39 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:187:

CTGATGCGTC GGATCATCCA GCAGCAGCAG CAGCAGCAG

39

## (2) INFORMATION FOR SEQ ID NO:188:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 40 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:188:

CTGATGCGTC GGATCATCCA GCAGCAGCAG CAGCAGCAGC

40

## (2) INFORMATION FOR SEQ ID NO:189:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 41 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

-297-

- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:189:

CTGATGCGTC GGATCATCCA GCAGCAGCAG CAGCAGCAGC A

41

(2) INFORMATION FOR SEQ ID NO:190:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 42 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:190:

CTGATGCGTC GGATCATCCA GCAGCAGCAG CAGCAGCAGC AG

42

(2) INFORMATION FOR SEQ ID NO:191:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 43 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:191:

CTGATGCGTC GGATCATCCA GCAGCAGCAG CAGCAGCAGC AGT

43

(2) INFORMATION FOR SEQ ID NO:192:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 44 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:192:

-298-

CTGATGCGTC GGATCATCCA GCAGCAGCAG CAGCAGCAGC AGTC

44

(2) INFORMATION FOR SEQ ID NO:193:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 45 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:193:

CTGATGCGTC GGATCATCCA GCAGCAGCAG CAGCAGCAGC AGTCA

45

(2) INFORMATION FOR SEQ ID NO:194:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 46 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:194:

CTGATGCGTC GGATCATCCA GCAGCAGCAG CAGCAGCAGC AGTCAC

46

(2) INFORMATION FOR SEQ ID NO:195:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 47 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:195:

CTGATGCGTC GGATCATCCA GCAGCAGCAG CAGCAGCAGC AGTCACG

47

(2) INFORMATION FOR SEQ ID NO:196:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 48 base pairs
- (B) TYPE: nucleic acid



-299-

- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:196:

CTGATGCGTC GGATCATCCA GCAGCAGCAG CAGCAGCAGC AGTCACGC

48

- (2) INFORMATION FOR SEQ ID NO:197:

- (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 49 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:197:

CTGATGCGTC GGATCATCCA GCAGCAGCAG CAGCAGCAGC AGTCACGCT

49

- (2) INFORMATION FOR SEQ ID NO:198:

- (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:198:

CTGATGCGTC GGATCATCCA GCAGCAGCAG CAGCAGCAGC AGTCACGCTA

50

- (2) INFORMATION FOR SEQ ID NO:199:

- (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 51 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:

-300-

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:199:

CTGATGCGTC GGATCATCCA GCAGCAGCAG CAGCAGCAGC AGTCACGCTA A

51

(2) INFORMATION FOR SEQ ID NO:200:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 52 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:200:

CTGATGCGTC GGATCATCCA GCAGCAGCAG CAGCAGCAGC AGTCACGCTA AC

52

(2) INFORMATION FOR SEQ ID NO:201:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 53 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:201:

CTGATGCGTC GGATCATCCA GCAGCAGCAG CAGCAGCAGC AGTCACGCTA ACC

53

(2) INFORMATION FOR SEQ ID NO:202:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 54 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:202:

CTGATGCGTC GGATCATCCA GCAGCAGCAG CAGCAGCAGC AGTCACGCTA ACCG

54

(2) INFORMATION FOR SEQ ID NO:203:

-301-

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 55 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

## (ii) MOLECULE TYPE: cDNA

## (iii) HYPOTHETICAL: NO

## (iv) ANTISENSE: NO

## (v) FRAGMENT TYPE:

## (vi) ORIGINAL SOURCE:

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:203:

CTGATGCGTC GGATCATCCA GCAGCAGCAG CAGCAGCAGC AGTCACGCTA ACCGA

55

## (2) INFORMATION FOR SEQ ID NO:204:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 56 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

## (ii) MOLECULE TYPE: cDNA

## (iii) HYPOTHETICAL: NO

## (iv) ANTISENSE: NO

## (v) FRAGMENT TYPE:

## (vi) ORIGINAL SOURCE:

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:204:

CTGATGCGTC GGATCATCCA GCAGCAGCAG CAGCAGCAGC AGTCACGCTA ACCGAA

56

## (2) INFORMATION FOR SEQ ID NO:205:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 57 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

## (ii) MOLECULE TYPE: cDNA

## (iii) HYPOTHETICAL: NO

## (iv) ANTISENSE: NO

## (v) FRAGMENT TYPE:

## (vi) ORIGINAL SOURCE:

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:205:

CTGATGCGTC GGATCATCCA GCAGCAGCAG CAGCAGCAGC AGTCACGCTA ACCGAAT

57

## (2) INFORMATION FOR SEQ ID NO:206:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 58 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

## (ii) MOLECULE TYPE: cDNA

-302-

- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:206:

CTGATGCGTC GGATCATCCA GCAGCAGCAG CAGCAGCAGC AGTCACGCTA ACCGAATC 58

(2) INFORMATION FOR SEQ ID NO:207:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 59 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:207:

CTGATGCGTC GGATCATCCA GCAGCAGCAG CAGCAGCAGC AGTCACGCTA ACCGAATCC 59

(2) INFORMATION FOR SEQ ID NO:208:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 60 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:208:

CTGATGCGTC GGATCATCCA GCAGCAGCAG CAGCAGCAGC AGTCACGCTA ACCGAATCCC 60

(2) INFORMATION FOR SEQ ID NO:209:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 61 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:209:

-303-

CTGATGCGTC GGATCATCCA GCAGCAGCAG CAGCAGCAGC AGTCACGCTA ACCGAATCCC 60  
T 61

## (2) INFORMATION FOR SEQ ID NO:210:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 62 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:210:

CTGATGCGTC GGATCATCCA GCAGCAGCAG CAGCAGCAGC AGTCACGCTA ACCGAATCCC 60  
TG 62

## (2) INFORMATION FOR SEQ ID NO:211:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 63 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:211:

CTGATGCGTC GGATCATCCA GCAGCAGCAG CAGCAGCAGC AGTCACGCTA ACCGAATCCC 60  
TGG 63

## (2) INFORMATION FOR SEQ ID NO:212:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 64 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:212:

CTGATGCGTC GGATCATCCA GCAGCAGCAG CAGCAGCAGC AGTCACGCTA ACCGAATCCC 60  
TGGT 64

## (2) INFORMATION FOR SEQ ID NO:213:

-304-

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 65 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

## (ii) MOLECULE TYPE: cDNA

## (iii) HYPOTHETICAL: NO

## (iv) ANTISENSE: NO

## (v) FRAGMENT TYPE:

## (vi) ORIGINAL SOURCE:

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:213:

CTGATGCGTC GGATCATCCA GCAGCAGCAG CAGCAGCAGC AGTCACGCTA ACCGAATCCC 60  
TGGTC 65

## (2) INFORMATION FOR SEQ ID NO:214:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 66 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

## (ii) MOLECULE TYPE: cDNA

## (iii) HYPOTHETICAL: NO

## (iv) ANTISENSE: NO

## (v) FRAGMENT TYPE:

## (vi) ORIGINAL SOURCE:

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:214:

CTGATGCGTC GGATCATCCA GCAGCAGCAG CAGCAGCAGC AGTCACGCTA ACCGAATCCC 60  
TGGTCA 66

## (2) INFORMATION FOR SEQ ID NO:215:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 67 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

## (ii) MOLECULE TYPE: cDNA

## (iii) HYPOTHETICAL: NO

## (iv) ANTISENSE: NO

## (v) FRAGMENT TYPE:

## (vi) ORIGINAL SOURCE:

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:215:

CTGATGCGTC GGATCATCCA GCAGCAGCAG CAGCAGCAGC AGTCACGCTA ACCGAATCCC 60  
TGGTCAG 67

## (2) INFORMATION FOR SEQ ID NO:216:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 68 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single

-305-

(D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:216:

CTGATGCGTC GGATCATCCA GCAGCAGCAG CAGCAGCAGC AGTCACGCTA ACCGAATCCC 60  
TGGTCAGA 68

(2) INFORMATION FOR SEQ ID NO:217:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 69 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:217:

CTGATGCGTC GGATCATCCA GCAGCAGCAG CAGCAGCAGC AGTCACGCTA ACCGAATCCC 60  
TGGTCAGAT 69

(2) INFORMATION FOR SEQ ID NO:218:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 70 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:218:

CTGATGCGTC GGATCATCCA GCAGCAGCAG CAGCAGCAGC AGTCACGCTA ACCGAATCCC 60  
TGGTCAGATC 70

(2) INFORMATION FOR SEQ ID NO:219:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 71 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO

-306-

- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:219:

CTGATGCGTC GGATCATCCA GCAGCAGCAG CAGCAGCAGC AGTCACGCTA ACCGAATCCC 60  
TGGTCAGATC T 71

(2) INFORMATION FOR SEQ ID NO:220:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 72 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:220:

CTGATGCGTC GGATCATCCA GCAGCAGCAG CAGCAGCAGC AGTCACGCTA ACCGAATCCC 60  
TGGTCAGATC TT 72

(2) INFORMATION FOR SEQ ID NO:221:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 13 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:221:

TGCACCTGAC TCC

13

(2) INFORMATION FOR SEQ ID NO:222:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 14 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:222:



-307-

TGCACCTGAC TCCT

14

(2) INFORMATION FOR SEQ ID NO:223:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 15 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:223:

TGCACCTGAC TCCTG

15

(2) INFORMATION FOR SEQ ID NO:224:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 16 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:224:

TGCACCTGAC TCCTGT

16

(2) INFORMATION FOR SEQ ID NO:225:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 17 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:225:

TGCACCTGAC TCCTGTG

17

(2) INFORMATION FOR SEQ ID NO:226:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 18 base pairs
- (B) TYPE: nucleic acid

-308-

- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:226:

TGCACCTGAC TCCTGTGG

18

(2) INFORMATION FOR SEQ ID NO:227:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 19 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:227:

TGCACCTGAC TCCTGTGGA

19

(2) INFORMATION FOR SEQ ID NO:228:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 20 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:228:

TGCACCTGAC TCCTGTGGAG

20

(2) INFORMATION FOR SEQ ID NO:229:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 21 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:

-309-

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:229:

TGCACCTGAC TCCTGTGGAG A

21

(2) INFORMATION FOR SEQ ID NO:230:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 22 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:230:

TGCACCTGAC TCCTGTGGAG AA

22

(2) INFORMATION FOR SEQ ID NO:231:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 23 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:231:

TGCACCTGAC TCCTGTGGAG AAG

23

(2) INFORMATION FOR SEQ ID NO:232:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 24 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:232:

TGCACCTGAC TCCTGTGGAG AAGT

24

(2) INFORMATION FOR SEQ ID NO:233:

-310-

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 25 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

## (ii) MOLECULE TYPE: cDNA

## (iii) HYPOTHETICAL: NO

## (iv) ANTISENSE: NO

## (v) FRAGMENT TYPE:

## (vi) ORIGINAL SOURCE:

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:233:

TGCACCTGAC TCCTGTGGAG AAGTC

25

## (2) INFORMATION FOR SEQ ID NO:234:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 26 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

## (ii) MOLECULE TYPE: cDNA

## (iii) HYPOTHETICAL: NO

## (iv) ANTISENSE: NO

## (v) FRAGMENT TYPE:

## (vi) ORIGINAL SOURCE:

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:234:

TGCACCTGAC TCCTGTGGAG AAGTCT

26

## (2) INFORMATION FOR SEQ ID NO:235:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 27 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

## (ii) MOLECULE TYPE: cDNA

## (iii) HYPOTHETICAL: NO

## (iv) ANTISENSE: NO

## (v) FRAGMENT TYPE:

## (vi) ORIGINAL SOURCE:

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:235:

TGCACCTGAC TCCTGTGGAG AAGTCTG

27

## (2) INFORMATION FOR SEQ ID NO:236:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 28 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

## (ii) MOLECULE TYPE: cDNA

-311-

- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:236:

TGCACCTGAC TCCTGTGGAG AAGTCTGC

28

(2) INFORMATION FOR SEQ ID NO:237:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 29 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:237:

TGCACCTGAC TCCTGTGGAG AAGTCTGCC

29

(2) INFORMATION FOR SEQ ID NO:238:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 30 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:238:

TGCACCTGAC TCCTGTGGAG AAGTCTGCCG

30

(2) INFORMATION FOR SEQ ID NO:239:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 31 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:239:

-312-

TGCACCTGAC TCCTGTGGAG AAGTCTGCCG T

31

(2) INFORMATION FOR SEQ ID NO:240:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 32 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:240:

TGCACCTGAC TCCTGTGGAG AAGTCTGCCG TT

32

(2) INFORMATION FOR SEQ ID NO:241:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 33 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:241:

TGCACCTGAC TCCTGTGGAG AAGTCTGCCG TTA

33

(2) INFORMATION FOR SEQ ID NO:242:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 34 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:242:

TGCACCTGAC TCCTGTGGAG AAGTCTGCCG TTAC

34

(2) INFORMATION FOR SEQ ID NO:243:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 35 base pairs
- (B) TYPE: nucleic acid

-313-

- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:243:

TGCACCTGAC TCCTGTGGAG AAGTCTGCCG TTACT

35

(2) INFORMATION FOR SEQ ID NO:244:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 36 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:244:

TGCACCTGAC TCCTGTGGAG AAGTCTGCCG TTACTG

36

(2) INFORMATION FOR SEQ ID NO:245:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 37 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:245:

TGCACCTGAC TCCTGTGGAG AAGTCTGCCG TTACTGC

37

(2) INFORMATION FOR SEQ ID NO:246:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 38 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:

-314-

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:246:

TGCACCTGAC TCCTGTGGAG AAGTCTGCCG TTA CTGCC

38

(2) INFORMATION FOR SEQ ID NO:247:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 39 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:247:

TGCACCTGAC TCCTGTGGAG AAGTCTGCCG TTA CTGCCC

39

(2) INFORMATION FOR SEQ ID NO:248:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 40 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:248:

TGCACCTGAC TCCTGTGGAG AAGTCTGCCG TTA CTGCCCT

40

(2) INFORMATION FOR SEQ ID NO:249:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 41 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:249:

TGCACCTGAC TCCTGTGGAG AAGTCTGCCG TTA CTGCCCT G

41

(2) INFORMATION FOR SEQ ID NO:250:



-315-

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 42 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

## (ii) MOLECULE TYPE: cDNA

## (iii) HYPOTHETICAL: NO

## (iv) ANTISENSE: NO

## (v) FRAGMENT TYPE:

## (vi) ORIGINAL SOURCE:

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:250:

TGCACCTGAC TCCTGTGGAG AAGTCTGCCG TTACTGCCCT GT

42

## (2) INFORMATION FOR SEQ ID NO:251:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 43 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

## (ii) MOLECULE TYPE: cDNA

## (iii) HYPOTHETICAL: NO

## (iv) ANTISENSE: NO

## (v) FRAGMENT TYPE:

## (vi) ORIGINAL SOURCE:

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:251:

TGCACCTGAC TCCTGTGGAG AAGTCTGCCG TTACTGCCCT GTG

43

## (2) INFORMATION FOR SEQ ID NO:252:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 44 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

## (ii) MOLECULE TYPE: cDNA

## (iii) HYPOTHETICAL: NO

## (iv) ANTISENSE: NO

## (v) FRAGMENT TYPE:

## (vi) ORIGINAL SOURCE:

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:252:

TGCACCTGAC TCCTGTGGAG AAGTCTGCCG TTACTGCCCT GTGG

44

## (2) INFORMATION FOR SEQ ID NO:253:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 45 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

## (ii) MOLECULE TYPE: cDNA

-316-

- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:253:

TGCACCTGAC TCCTGTGGAG AAGTCTGCCG TTACTGCCCT GTGGG

45

(2) INFORMATION FOR SEQ ID NO:254:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 46 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:254:

TGCACCTGAC TCCTGTGGAG AAGTCTGCCG TTACTGCCCT GTGGGG

46

(2) INFORMATION FOR SEQ ID NO:255:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 47 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:255:

TGCACCTGAC TCCTGTGGAG AAGTCTGCCG TTACTGCCCT GTGGGGC

47

(2) INFORMATION FOR SEQ ID NO:256:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 48 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:256:

-317-

TGCACCTGAC TCCTGTGGAG AAGTCTGCCG TTAGTCCCCT GTGGGGCA

48

## (2) INFORMATION FOR SEQ ID NO:257:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 49 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:257:

TGCACCTGAC TCCTGTGGAG AAGTCTGCCG TTAGTCCCCT GTGGGGCAA

49

## (2) INFORMATION FOR SEQ ID NO:258:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:258:

TGCACCTGAC TCCTGTGGAG AAGTCTGCCG TTAGTCCCCT GTGGGGCAAG

50

## (2) INFORMATION FOR SEQ ID NO:259:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 51 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:259:

TGCACCTGAC TCCTGTGGAG AAGTCTGCCG TTAGTCCCCT GTGGGGCAAG G

51

## (2) INFORMATION FOR SEQ ID NO:260:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 52 base pairs
- (B) TYPE: nucleic acid

-318-

- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:260:

TGCACCTGAC TCCTGTGGAG AAGTCTGCCG TTACTGCCCT GTGGGGCAAG GT 52

(2) INFORMATION FOR SEQ ID NO:261:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 209 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:261:

CATTTGCTTC TGACACAAC GTGTTCACTA GCAACCTCAA ACAGACACCA TGGTGCACCT 60  
GACTCCTGTG GAGAAGTCTG CCGTTACTGC CCTGTGGGGC AAGGTGAACG TGGATGAAAT 120  
TGGTGGTGAG GCCCTGGGCA GGTGGTATC AAGGTTACAA GACAGGTTTA AGGAGACCAA 180  
TAGAAACTGG GCATGTGGAG ACAGAGAAG 209

(2) INFORMATION FOR SEQ ID NO:262:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 88 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:262:

TGAGACTCTG TCTCAAAAAT AAATAAATAA ATAAATAAAT AAATAAATAA ATAAATAAAT 60  
AAATAAATAA GTAAAAAGA AAGAAATGC 88

(2) INFORMATION FOR SEQ ID NO:263:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 54 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: unknown

-319-

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:263:

GTGTGTGTGT GTGTGTGTT TTTTTTAAAC AGGGATTG GGAATTATTT GAGA

54

(2) INFORMATION FOR SEQ ID NO:264:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 24 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:264:

TTCCCCAAAT CCCTGTATAA AAC

23

(2) INFORMATION FOR SEQ ID NO:265:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 26 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:265:

TTCCCCAAAT CCCTGTATAA AAAAC

25

(2) INFORMATION FOR SEQ ID NO:266:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 27 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

-320-

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:266:

TTCCCCAAAT CCCTGTTAAA AAAAAAC

27

(2) INFORMATION FOR SEQ ID NO:267:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 103 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:267:

GTAAAACGAC CGCCAGTGCC AAGCTTGCAAT GCCTGCAGGT CGACTCTAGA GGATCCCCGG 60  
GTACCGAGCT CGAATTCGTA ATCATGGTCA TAGCTGTTTC CTG 103

(2) INFORMATION FOR SEQ ID NO:268:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 78 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:268:

GAGTCAGGTG CGCCATGCCT CAAACAGACA CCATGGTGCA CCTGACTCCT GAGGAGNCTG 60  
GGCATGTGGA GACAGAGA 78

(2) INFORMATION FOR SEQ ID NO:269:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 78 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:269:

TCTCTGTCTC CACATGCCCA GNCTCCTCAG GACTCAGGTG CACATGGTGT CTGTTTGAGG 60

-321-

CATGGCGCAC CTGAGCTC

78

## (2) INFORMATION FOR SEQ ID NO:270:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 78 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:270:

TCTCTGTCTC CACATGCCCA GNCTCCTCAG GAGTCAGGTG CGCCATGGTG TCTGTTTGAG 60  
GCATGGCGCA CGTGACTC 78

## (2) INFORMATION FOR SEQ ID NO:271:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 82 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:271:

TCTCTGTCTC CACATGCCCA GNCTCCTCAG GAGTCAGGTG CGCCATGGTG TCTGTTTGAG 60  
GCATGGCGCA CCTGACTCCT GA 82

## (2) INFORMATION FOR SEQ ID NO:272:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 42 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:272:

TCTCTGTCTC CACATGCCCA GNCTCCTCAG GAGTCAGGTG CG 42

## (2) INFORMATION FOR SEQ ID NO:273:

## (i) SEQUENCE CHARACTERISTICS:

-322-

- (A) LENGTH: 13 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:273:

CACCTGACTC CTA

13

(2) INFORMATION FOR SEQ ID NO:274:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 14 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:274:

CACCTGACTC CTGGA

14

(2) INFORMATION FOR SEQ ID NO:275:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 14 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:275:

CACCTGACTC CTGA

14

(2) INFORMATION FOR SEQ ID NO:276:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 26 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO



-323-

- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:276:

CCATGGTGTC TGTTTGAGGC ATGGCG

26

(2) INFORMATION FOR SEQ ID NO:277:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 75 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:277:

CAGCTCTCAT TTTCCATACA GTCAGTATCA ATTCTGGAAG AATTTCAGAG CATTAAAGAT 60  
AGTCATCTTG GGGCT 75

(2) INFORMATION FOR SEQ ID NO:278:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 61 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:278:

ACCTAGCGTT CAGTTCGACT GAGATAATAC GACTCACTAT AGCAGCTCTC ATTTCCATA 60  
C 61

(2) INFORMATION FOR SEQ ID NO:279:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 20 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: RNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:279:

-324-

GUCACUACAG GUGAGCUCCA

20

## (2) INFORMATION FOR SEQ ID NO:280:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 24 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:280:

CTCAGTCCAC GTGGTACCCT GCTG

24

## (2) INFORMATION FOR SEQ ID NO:281:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 85 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:281:

CATTTGCTTC TGACACAAC GTGTTCACTA GCAACCTCAA ACAGACACCA TGGTGACCT  
GACTCCTGAG GAGAAGTCTG CCGTT

85

## (2) INFORMATION FOR SEQ ID NO:282:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 22 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:282:

ACGGGTCCCG GAGTGGTCTC GC

22

## (2) INFORMATION FOR SEQ ID NO:283:

-325-

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 76 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:283:

ACTGCCCTGT GGGGCAAGGT GAACGTGGAT GAAGTTGGTG GTGAGGCCCT GGGCAGGTTG 60  
GTATCAAGGT TACAAG 76

## (2) INFORMATION FOR SEQ ID NO:284:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 76 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:284:

ACTGCCCTGT GGGGCAAGGT GAACGTGGAT GAAGTTGGTG GTGAGGCCCT GGGCAGATTG 60  
GTATCAAGGT TACAAG 76

## (2) INFORMATION FOR SEQ ID NO:285:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 76 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:285:

ACTGCCCTGT GGGGCAAGGT GAACGTGGAT GAAGTTGGTG GTGAGGCCCT GGGCAGGTTG 60  
GTATCAAGGT TACAAG 76

## (2) INFORMATION FOR SEQ ID NO:286:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 76 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single

-326-

(D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:286:

ACTGCCCTGT GGGGCAAGGT GAACGTGGAT GAAGTTGGTG GTGAGGCCCT GGGCAGGTG 60  
GCATCAAGGT TACAAG 76

(2) INFORMATION FOR SEQ ID NO:287:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 48 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:287:

ACAGGTTTAA GGAGACCAAT AGAACTGGG CATGTGGAGA CAGAGAAG 48

(2) INFORMATION FOR SEQ ID NO:288:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 24 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:288:

GACGACGACT GCTACCTGAC TCCA 24

(2) INFORMATION FOR SEQ ID NO:289:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 24 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:

-327-

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:289:

ACAGCGCACT GCTACCTGAC TCCA

24

(2) INFORMATION FOR SEQ ID NO:290:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 18 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:290:

TGGAGTCAGG TAGCAGTC

18

(2) INFORMATION FOR SEQ ID NO:291:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 60 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:291:

CAGCTCTCAT TTTCCATACA GTCAGTATCA ATTCTGGAAG AATTTCCAGA CATTAAAGAT 60

(2) INFORMATION FOR SEQ ID NO:292:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 65 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:292:

AGTCATCTTG GGGCTGTGCA GAGTAAAAGG TATGTCAGTC ATAGTTAAGA CCTTCTTAAA  
GGTCT

65

-328-

## (2) INFORMATION FOR SEQ ID NO:293:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 25 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:293:

GTAATTCTA TCAGTAGAAC CCCGA

25

## (2) INFORMATION FOR SEQ ID NO:294:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 60 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:294:

CAGCTCTCAT TTTCCATACA GTCAGTATCA ATTCTGGAAG AATTCCAGA CATTAAAGAT 60

## (2) INFORMATION FOR SEQ ID NO:295:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 15 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:295:

AGTCATCTTG GGGCT

15

## (2) INFORMATION FOR SEQ ID NO:296:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 60 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

-329-

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:296:

CAGCTCTCAT TTTCCATACA GTCAGTATCA ATTCTGGAAG AATTTCCAGA CATTAAAGAT 60

(2) INFORMATION FOR SEQ ID NO:297:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 16 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:297:

AGTCATCTTG GGGCTA

16

(2) INFORMATION FOR SEQ ID NO:298:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 43 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:298:

CAGCTCTCAT TTTCCATACA TTAAAGATAG TCATCTTGGG GCT

43

(2) INFORMATION FOR SEQ ID NO:299:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 44 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:299:

-330-

CAGCTCTCAT TTTCCATACA TTAAAGATAG TCATCTGGG GCTA

44

## (2) INFORMATION FOR SEQ ID NO:300:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 22 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:300:

CAGCTCTCAT TTTCCATACA GT

22

## (2) INFORMATION FOR SEQ ID NO:301:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 21 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:301:

CAGCTCTCAT TTTCCATACA T

21

## (2) INFORMATION FOR SEQ ID NO:302:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:302:

GCCTGGTACA CTGCCAGGCG CTTCTGCAGG TCATCGGCAT CGCGGAGGAG

50

## (2) INFORMATION FOR SEQ ID NO:303:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid



-331-

- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:303:

GCCTGGTACA CTGCCAGGCA CTTCTGCAGG TCATCGGCAT CGCGGAGGAG

50

(2) INFORMATION FOR SEQ ID NO:304:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 21 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:304:

GATGCCGATG ACCTGCAGAA G

21

(2) INFORMATION FOR SEQ ID NO:305:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 22 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:305:

GATGCCGATG ACCTGCAGAA GC

22

(2) INFORMATION FOR SEQ ID NO:306:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 24 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO

-332-

- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:306:

GATGCCGATG ACCTGCAGAA GTGC

24

(2) INFORMATION FOR SEQ ID NO:307:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 12 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:307:

GATGATCCGA CG

12

(2) INFORMATION FOR SEQ ID NO:308:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 18 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:308:

CTGATGCGTC GGATCATC

18

(2) INFORMATION FOR SEQ ID NO:309:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 12 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:309:

GATGATCCGA CG

12

-333-

## (2) INFORMATION FOR SEQ ID NO:310:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 35 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:310:

GGCGCGGACA TGGAGGACGT GTGCGGCCGC CTGGT

35

## (2) INFORMATION FOR SEQ ID NO:311:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 34 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:311:

TCCGCGATGC CGATGACCTG CAGAAGCGCC TGGC

34

## (2) INFORMATION FOR SEQ ID NO:312:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 27 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:312:

CGGCTGCGAT CACCGTGCGG CACAGCT

27

## (2) INFORMATION FOR SEQ ID NO:313:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 27 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single

-334-

(D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:313:

CGGCTGCGAT CACCGTGCGG T

21

(2) INFORMATION FOR SEQ ID NO:314:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 27 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:314:

CGGCTGCGAT CACCGTGCGG AACAGCT

27

(2) INFORMATION FOR SEQ ID NO:315:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 22 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:315:

CGGCTGCGAT CACCGTGCGG CA

22

(2) INFORMATION FOR SEQ ID NO:316:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 22 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

-335-

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:316:

CGGCTGCGAT CACCGTGCGG TA

22

(2) INFORMATION FOR SEQ ID NO:317:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 21 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:317:

CGGCTGCGAT CACCGTGCGG A

21

(2) INFORMATION FOR SEQ ID NO:318:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 42 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:318:

ATCATCAACT GGAAGATCAG GTCAGGAGCC ACTTGCCANC CT

42

(2) INFORMATION FOR SEQ ID NO:319:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 33 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:319:

ATCATCACAC TGGAAGACTC CAGGTCAGGA GCC

33

(2) INFORMATION FOR SEQ ID NO:320:

(i) SEQUENCE CHARACTERISTICS:

-336-

- (A) LENGTH: 48 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:320:

ATCCACTACA ACTACATGTG TAACAGTTGG wGCwwGCC

48

-337-

**WHAT IS CLAIMED IS:**

1. A process for determining the sequence of a target nucleic acid molecule comprising the steps of:

a) generating at least two nucleic acid fragments from the target nucleic acid; and

b) analyzing the at least two fragments by a mass spectrometry format, and thereby determine the sequence of the target nucleic acid molecule.

2. A process of claim 1, wherein in step a), an endonuclease is contacted with the target nucleic acid to generate the at least two nucleic acid fragments.

3. A process of claim 2, wherein the endonuclease is a restriction enzyme that can recognize and cleave at least one restriction site in the target nucleic acid.

4. A process of claim 2, wherein the target nucleic acid is a deoxyribonucleic acid and the nuclease is a deoxyribonuclease.

5. A process of claim 2, wherein the target nucleic acid is a ribonucleic acid and the nuclease is a ribonuclease.

6. A process of claim 5, wherein the ribonuclease is selected from the group consisting of: the G-specific T<sub>1</sub> ribonuclease, the A-specific U<sub>2</sub> ribonuclease, the A/U specific PhyM ribonuclease, the U/C specific ribonuclease A, the C-specific chicken liver ribonuclease and crisavirin.

7. A process of claim 1, wherein in step a), nucleic acid fragments are generated by performance of a combined amplification and base-specific termination reaction.

8. A process of claim 7, wherein the combined amplification and base-specific termination reaction is performed using a first polymerase, which has a relatively low affinity towards at least one chain terminating

-338-

nucleotide and an a second polymerase, which has a relatively high affinity towards at least one chain terminating nucleotide.

9. A process of claim 8, wherein the first and second polymerases are thermostable DNA polymerases.

5. 10. A process of claim 9, wherein the thermostable DNA polymerases are selected from the group consisting of: Taq DNA polymerase, AmpliTaq FS DNA polymerase, Deep Vent (exo-) DNA polymerase, Vent DNA polymerase, Vent (exo') DNA polymerase, Vent DNA polymerase, Vent (exo') DNA polymerase, Deep Vent DNA polymerase, Thermo Sequenase, *exo(-) Pseudococcus furiosus (Pfu)* DNA polymerase, AmpliTaq, Ultman, 9 degree Nm, Tth, Hot Tub, *Pyrococcus furiosus (Pfu)* and *Pyrococcus woesei (Pwo)* DNA polymerase.

11. A process of claim 1, wherein the at least two nucleic acid fragments generated in step a) include mass modified nucleotides.

15. 12. A process of claim 1, wherein the at least two fragments comprise a 3' tag.

13. A process of claim 1, wherein the at least two fragments comprise a 5' tag.

20. 14. A process of claim 12 or 13, wherein the tag is a non-natural tag.

15. A process of claim 14, wherein the non-natural tag is selected from the group consisting of: an affinity tag and a mass marker.

16. A process of claim 15, wherein the affinity tag facilitates immobilization of the nucleic acid to a solid support.

25. 17. A process of claim 16, wherein the affinity tag is biotin or a nucleic acid sequence that is capable of binding to a capture nucleic acid sequence that is bound to a solid support.



-339-

18. A process of claim 1, wherein the process additionally comprises the step of: ordering the at least two nucleic acid fragments to determine the sequence of the target nucleic acid.

19. A process for detecting a target nucleic acid present in a biological sample, comprising the steps of:

5 a) performing on a nucleic acid obtained from a biological sample; a first polymerase chain reaction using a first set of primers, which are capable of amplifying a portion of the nucleic acid containing the target nucleic acid, thereby producing a first amplification product; and

10 b) detecting the first amplification product by mass spectrometry, wherein detection of the target nucleic acid indicates that the target nucleic acid is present in the biological sample.

20. A process of claim 19, wherein prior to step b), a second polymerase chain reaction is performed on the first amplification product using a second set of primers, which are capable of amplifying at least a portion of the first amplification product, which contains the target nucleic acid.

21. A process of claim 19 or 20, wherein prior to step b), the target nucleic acid is immobilized to a solid support.

22. A process of claim 21, wherein the target nucleic acid is reversibly immobilized.

23. A process of claim 22, wherein the target nucleic acid can be cleaved from the solid support by a chemical, enzymatic or physical process.

24. A process of claim 23, wherein immobilization is accomplished via a photocleavable bond.

25. A process of claim 22, wherein the target nucleic acid is cleaved from the support during step b).

-340-

26. A process of claim 21, wherein the solid support is selected from the group consisting of: beads, flat surfaces, chips, capillaries, pins, combs and wafers.

27. A process of claim 21, wherein immobilization is  
5 accomplished by hybridization between a complementary capture nucleic acid molecule immobilized to a solid support, and a portion of the nucleic acid molecule, which is distinct from the target nucleic acid sequence.

28. A process of claim 19 or 20, wherein prior to step b), the target nucleic acid is purified.

10 29. A process of claim 19 or 20, wherein the primer or first or second amplification product is conditioned.

30. A process of claim 29, wherein the primer or first or second amplification product is conditioned by phosphodiester backbone modification.

15 31. A process of claim 30, wherein the phosphodiester backbone modification is a cation exchange.

32. A process of claim 29, wherein the primer or first or second amplification product is conditioned by contact with an alkylating agent or trialkylsilyl chloride.

20 33. A process of claim 29, wherein conditioning is effected by including at least one nucleotide that reduces sensitivity for depurination in the primer or first or second amplification product.

34. A process of claim 33, wherein the nucleotide is an N7- or N9- deazapurine nucleotide or 2' fluoro 2' deoxy nucleotide.

25 35. A method for detecting neoplasia/malignancies in a tissue or cell sample, comprising detecting telomerase activity, mutation of a proto-oncogene, expression of a tumor specific gene in the sample by detecting nucleic acids that encode the telomerase, that are specific for the mutation or that encode the tumor-specific by mass spectrometry.

-341-

36. The method of claim 35 that is a method for detecting neoplasia/malignancies in a tissue or cell sample, comprising:

- 5
- a) isolating telomerase from the sample and adding a synthetic DNA primer, which is optionally immobilized, complementary to a telomeric repeat, and all four deoxynucleotide triphosphates under conditions that result in telomerase specific extension of the synthetic DNA;
  - b) amplifying the telomerase extended DNA product; and
  - 10 c) detecting the DNA product by mass spectrometry, wherein telomerase-specific extension is indicative of neoplasia/malignancy.

37. The method of claim 36, wherein the primer contains a linker moiety for immobilization on a support; and the amplified primers are isolated conjugating the linker portion to a solid support.

15

38. The method of claim 35 that is a method for identifying transformed cells or tissues, comprising:

- 20
- a) in a cell or tissue sample, amplifying a portion of a proto-oncogene that includes a codon indicative of transformation, wherein one primer comprises a linker moiety for immobilization;
  - c) immobilizing DNA via the linker moiety to a solid support, optionally in the form of an array;
  - 25 d) hybridizing a primer complementary to the proto oncogene sequence that is upstream from the codon
  - e) adding 3dNTPs/1 ddNTP and DNA polymerase and extending the hybridized primer to the next ddNTP location;

-342-

- f) ionizing/volatizing the sample; and
- g) detecting the mass of the extended DNA, whereby mass indicates the presence of wild-type or mutant alleles. The presence of a mutant allele at the codon is diagnostic for neoplasia.

39. The method of claim 38, wherein the proto-oncogene is the RET-proto-oncogene.

40. The method of claim 35 that is a method for detecting expression of a tumor-specific gene, comprising:

- a) isolating polyA RNA from the sample;
- c) preparing a cDNA library using reverse transcription;
- d) amplifying a cDNA product, or portion thereof, of the tumor-specific gene, wherein one oligo primer comprises a linker moiety;
- e) isolating the amplified product by immobilizing the DNA to a solid support via the linker moiety;
- f) optionally conditioning the DNA;
- g) ionizing/volatizing sample and detecting the presence of a DNA peak that is indicative of expression of the gene.

41. The method of claim 40, wherein the cells are bone marrow cells, the gene is the tyrosine hydroxylase gene, and expression of the gene is indicative of neuroblastoma.

42. A method for directly detecting a double-stranded nucleic acid using matrix-assisted laser desorption/ionization (MALDI)-time-of-flight (TOF) mass spectrometry, comprising:

- a) isolating a double-stranded DNA fragment from a cell or tissue sample;

-343-

- b) preparing the double-stranded DNA for analysis under conditions that increase the ratio of dsDNA:ssDNA, wherein the conditions include one or all of the following: preparing samples for analysis at reduced temperatures (i.e. 4 ° C), and using of higher DNA concentrations in the matrix to drive duplex formation;
- c) ionizing/volatizing the sample of step b), wherein low acceleration voltage of the ions are used;
- d) detecting the presence of the double-stranded DNA.

5

43. A method for comparing DNA samples to discern relatedness or to detect mutations, comprising:

10

- a) obtaining biological a plurality of samples;
- b) amplifying a region of DNA from each sample that contains two or more microsatellite DNA repeat sequences;
- c) ionizing/volatizing the amplified DNA;
- d) detecting the presence of the amplified DNA and comparing the molecular weight of the amplified DNA, wherein different sizes are indicative of non-identity between or among the samples.

15

44. The method of claim 43, wherein non-identity is indicative of the presence of a mutation in the DNA in one sample, non-relatedness or non-HLA compatibility between or among the individuals from whom the samples were obtained.

20

45. The method of claim 43 or 44, wherein a plurality of markers are examined simultaneously.

25

-345-

47. A method for detecting a target nucleic acid in a biological sample using RNA amplification, comprising:

amplifying the target nucleic acid using a primer comprising a region complementary to the target sequence and a region that encodes

5 a promoter;

synthesizing RNA using an RNA polymerase that recognizes the promoter;

detecting the resulting RNA using mass spectrometry.

48. A primer for mass spectrometric analyses, comprising all or  
10 at least about 20, preferably about 16, bases of any of the sequence of nucleotide sequences set forth in SEQ ID NOs. 1-22, 24, 27-38, 41-86, 89, 92, 95, 98, 101-110, 112-123, 126, 128 and 129, wherein the primer is unlabeled.

49. The primer of claim 48, further comprising a mass  
15 modifying moiety.

50. A process for detecting a target nucleic acid sequence present in a biological sample, comprising the steps of:

a) obtaining a nucleic acid molecule containing a target nucleic acid sequence from a biological sample;

20 b) immobilizing the target sequence on the support via thiol linkages, whereby the target is present at a sufficient density to detect it using mass spectrometry;

c) hybridizing a detector oligonucleotide with the target nucleic acid sequence;

25 d) removing unhybridized detector oligonucleotide;

e) ionizing and volatilizing the product of step c); and

f) detecting the detector oligonucleotide by mass spectrometry, wherein detection of the detector oligonucleotide indicates the presence of the target nucleic acid sequence in the biological sample.

-346-

51. The process of claim 50, wherein the target nucleic acid molecule is amplified prior to immobilization.

52. The process of claim 50 or 52, wherein at least one of the detector oligonucleotide or the target nucleic acid sequence has been  
- 5 - conditioned.

53. A process of any of claims 50-52, wherein the solid support is selected from the group consisting of: beads, flat surfaces, pins and combs.

54. A process of any of claims 50-53, wherein target nucleic  
10 acid is immobilized in the form of an array.

55. A process of any of claims 50-54, wherein the support is a silicon wafer.

56. A process of any of claims 51-55, wherein the target nucleic acid molecule is amplified by an amplification procedure selected  
15 from the group consisting of cloning, transcription, the polymerase chain reaction (PCR), the ligase chain reaction (LCR), and strand displacement amplification (SDA).

57. A process of any of claims 50-56, wherein the mass spectrometer is selected from the group consisting of: Matrix-Assisted  
20 Laser Desorption/Ionization Time-of-Flight (MALDI-TOF), Electrospray (ES), Ion Cyclotron Resonance (ICR), and Fourier Transform.

58. A process of any of claims 50-57, wherein the sample is conditioned by mass differentiating at least two detector oligonucleotides or oligonucleotide mimetics to detect and distinguish at least two target  
25 nucleic acid sequences simultaneously.

59. A process of claim 58, wherein the mass differentiation is achieved by differences in the length or sequence of the at least two oligonucleotides.

-347-

60. A process of claim 59, wherein the mass differentiation is achieved by the introduction of mass modifying functionalities in the base, sugar or phosphate moiety of the detector oligonucleotides.

61. A process of claim 58, wherein the mass differentiation is achieved by exchange of cations at the phosphodiester bond.

62. A process of any of claims 50-61, wherein the nucleic acid molecule obtained from a biological sample is amplified into DNA using mass modified dideoxynucleoside triphosphates and DNA dependent DNA polymerase prior to mass spectrometric detection.

63. A process of any of claims 50-62, wherein the nucleic acid molecule obtained from a biological sample is amplified into RNA using mass modified ribonucleoside triphosphates and DNA dependent RNA polymerase prior to mass spectrometric detection.

64. A process of any of claims 50-63, wherein the target nucleic acid sequence is indicative of a disease or condition selected from the group consisting of a genetic disease, a chromosomal abnormality, a genetic predisposition, a viral infection, a fungal infection and a bacterial infection.

65. A method of determining a sequence of a nucleic acid, comprising the steps of:

- (i) obtaining multiple copies of the nucleic acid to be sequenced;
- (ii) cleaving the multiple copies from a first end to a second end with an exonuclease to sequentially release individual nucleotides;
- (iii) identifying each of the sequentially released nucleotides by

mass spectrometry; and

- (iv) determining the sequence of the nucleic acid from the identified nucleotides, wherein the nucleic acid is immobilized by covalent attachment to a solid support via at least one sulfur atom.



-348-

66. A method of determining a sequence of a nucleic acid, comprising the steps of:

- (i) obtaining multiple copies of the nucleic acid to be sequenced;
- (ii) cleaving the multiple copies from a first end to a second end

5 with an exonuclease to produce multiple sets of nested nucleic acid fragments;

(iii) determining the molecular weight value of each one of the sets of nucleic acid fragments by mass spectrometry; and

(iv) determining the sequence of the nucleic acid from the  
10 molecular weight values of the sets of nucleic acid fragments, wherein the nucleic acid is immobilized by covalent attachment to a solid support via at least one sulfur atom.

67. The process of claim 65 or 66, wherein the nucleic acids are covalently bound to a surface of the support at a density of at least 20  
15 fmol/mm<sup>2</sup>.

68. The method of any of claims 50-67, wherein immobilization is effected by a method comprising:

reacting a thiol-containing insoluble support with a nucleic acid comprising a thiol-reactive group under conditions such that a covalent  
20 bond is formed;

thereby immobilizing the nucleic acid on the insoluble support.

69. The method of claim 68, further including the step of modifying the insoluble support with a thiol-containing reagent, to form a thiol-containing insoluble support.

25 70. The method of claim 68 or 69, wherein the thiol-reactive cross-linking reagent is N-succinimidyl (4-iodoacetyl) aminobenzoate (SIAB).

71. The method of claim 65 or claim 66, wherein the nucleic acid is a 2'-deoxyribonucleic acid (DNA).

-349-

72. The method of claim 65 or claim 66, wherein the nucleic acid is a ribonucleic acid (RNA).

73. The method of any of claims 65-71, wherein the exonuclease is selected from the group consisting of snake venom phosphodiesterase, spleen phosphodiesterase, Bal-31 nuclease, *E. coli* exonuclease I, *E. coli* exonuclease VII, Mung Bean Nuclease, S1 Nuclease, an exonuclease activity of *E. coli* DNA polymerase 1, an exonuclease activity of a Klenow fragment of DNA polymerase 1, an exonuclease activity of T4 DNA polymerase, an exonuclease activity of T7 DNA polymerase, an exonuclease activity of Taq DNA polymerase, an exonuclease activity of DEEP VENT DNA polymerase, *E. coli* exonuclease III, lambda exonuclease and an exonuclease activity of VENT<sub>R</sub> DNA polymerase.

74. The method of any of claims 65-74, wherein the nucleic acid comprises mass-modified nucleotides.

75. The method of claim 74, wherein the mass-modified nucleotides modulate the rate of the exonuclease activity.

76. The method of claim 74, wherein the sequentially released nucleotides are mass-modified subsequent to exonuclease release and prior to mass spectrometric identification.

77. The method of claim 76, wherein the sequentially released nucleotides are mass-modified by contact with an alkaline phosphatase.

78. A method of any of claims 65-77, wherein the mass spectrometry format is matrix assisted laser desorption (MALDI) mass spectrometry or electrospray (ES) mass spectrometry.

79. A method of any of claims 65-79, wherein immobilization is effected by a method, comprising:

-350-

reacting the surface of the substrate with a solution of 3-aminopropyltriethoxysilane to produce a uniform layer of primary amines on the surface of the substrate; and

derivatizing the surface of a substrate with iodoacetamido

5 functionalities by reacting the uniform layer of primary amines with a solution of N-succinimidyl (4-iodoacetyl) aminobenzoate (SIAB).

80. A primer, comprising all least about 20, preferably about 16, bases of any of the sequence of nucleotides sequences set forth in SEQ ID NOs. 1-22, 24, 27-38, 41-86, 89, 92, 95, 98, 101-110, 112-123,  
10 126, 128 and 129.

81. The primers of claim 80 that is unlabeled, and optionally includes a mass modifying moiety, which is preferably attached to the 5' end.

82. The method of any of claims 1-79, wherein nucleic acid is  
15 immobilized to a solid support via a selectively cleavable linker.

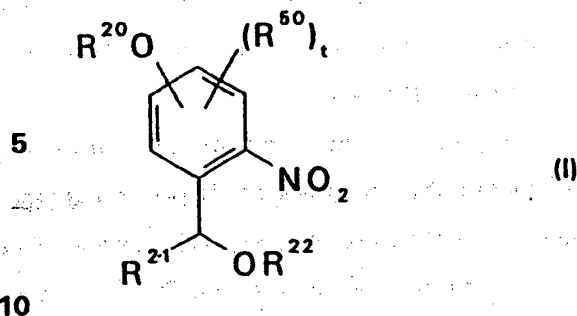
83. The method of claim 82, wherein the linker is thermocleavable, enzymatically cleavable, photocleavable or chemically cleavable.

82. The method of claim 82, wherein the linker is a trityl linker.

20 83. The method of claim 82, wherein the linker is selected from the group consisting of 1-(2-nitro-5-(3-O-4,4'-dimethoxytritylpropoxy)-phenyl)-1-O-((2-cyanoethoxy)-diisopropylaminophosphino)ethane and 1-(4-(3-O-4,4'-dimethoxytritylpropoxy)-3-methoxy-6-nitrophenyl)-1-O-((2-cyanoethoxy)-diisopropylaminophosphino)ethane.

25 84. A photolabile linker, comprising a compound of formula:

-351-



wherein:

$R^{20}$  is selected from the group consisting of  $\omega$ -(4,4'-dimethoxytrityloxy)alkyl and  $\omega$ -hydroxyalkyl;

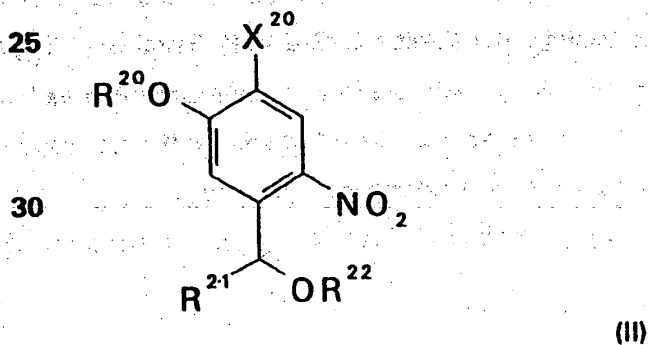
15  $R^{21}$  is selected from the group consisting of hydrogen, alkyl, aryl, alkoxycarbonyl, aryloxy carbonyl and carboxy;

$R^{22}$  is selected from the group consisting of hydrogen and (dialkylamino)( $\omega$ -cyanoalkoxy)P-;

$t$  is 0-3; and

20  $R^{50}$  is selected from the group consisting of alkyl, alkoxy, aryl and aryloxy.

85. The photocleavable linker of claim 84, wherein the linkers are of formula II:



-352-

wherein:

R<sup>20</sup> is selected from the group consisting of  $\omega$ -(4,4'-dimethoxytrityloxy)alkyl,  $\omega$ -hydroxyalkyl and alkyl;

5 R<sup>21</sup> is selected from the group consisting of hydrogen, alkyl, aryl, alkoxy carbonyl, aryloxy carbonyl and carboxy;

R<sup>22</sup> is selected from the group consisting of hydrogen and (dialkylamino)( $\omega$ -cyanoalkoxy)P-; and

10 X<sup>20</sup> is selected from the group consisting of hydrogen, alkyl or OR<sup>20</sup>.

86. The photocleavable linker of claim 85, wherein:

R<sup>20</sup> is selected from the group consisting of 3-(4,4'-dimethoxytrityloxy)propyl, 3-hydroxypropyl and methyl;

15 R<sup>21</sup> is selected from the group consisting of hydrogen, methyl and carboxy;

R<sup>22</sup> is selected from the group consisting of hydrogen and (diisopropylamino)(2-cyanoethoxy)P-; and

X<sup>20</sup> is selected from the group consisting of hydrogen, methyl or OR<sup>20</sup>.

20 87. The photocleavable linker of claim 85, wherein:

R<sup>20</sup> is 3-(4,4'-dimethoxytrityloxy)propyl;

R<sup>21</sup> is methyl;

R<sup>22</sup> is (diisopropylamino)(2-cyanoethoxy)P-; and

X<sup>20</sup> is hydrogen.

25 88. The photocleavable linker of claim 86, wherein:

R<sup>20</sup> is methyl;

R<sup>21</sup> is methyl;

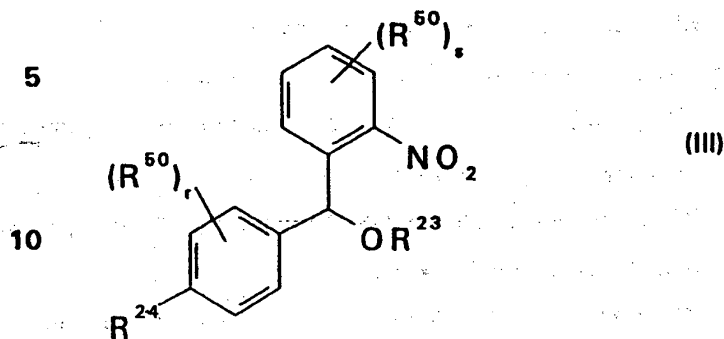
R<sup>22</sup> is (diisopropylamino)(2-cyanoethoxy)P-; and

X<sup>20</sup> is 3-(4,4'-dimethoxytrityloxy)propoxy.

-353-

88. A photocleavable linker, comprising a compound of formula

III:



15 wherein:

$R^{23}$  is selected from the group consisting of hydrogen and (dialkylamino)( $\omega$ -cyanoalkoxy)P-;

$R^{24}$  is selected from  $\omega$ -hydroxyalkoxy,  $\omega$ -(4,4'-dimethoxytrityloxy)alkoxy,  $\omega$ -hydroxyalkyl and  $\omega$ -(4,4'-

20 dimethoxytrityloxy)alkyl, and is unsubstituted or substituted on the alkyl or alkoxy chain with one or more alkyl groups;

r and s are each independently 0-4; and

$R^{50}$  is alkyl, alkoxy, aryl or aryloxy.

89. The photocleavable linker of claim 88, wherein:

25  $R^{24}$  is  $\omega$ -hydroxyalkyl or  $\omega$ -(4,4'-dimethoxytrityloxy)alkyl, and is substituted on the alkyl chain with a methyl group.

90. The photocleavable linker of claim 88, wherein:

$R^{23}$  is selected from the group consisting of hydrogen and (diisopropylamino)(2-cyanoethoxy)P-; and

30  $R^{24}$  is selected from the group consisting of 3-hydroxypropoxy, 3-(4,4'-dimethoxytrityloxy)propoxy, 4-hydroxybutyl, 3-hydroxy-1-propyl, 1-hydroxy-2-propyl, 3-hydroxy-2-methyl-1-propyl, 2-hydroxyethyl, hydroxymethyl, 4-(4,4'-dimethoxytrityloxy)butyl, 3-(4,4'-

-354-

dimethoxytrityloxy)-1-propyl, 2-(4,4'-dimethoxytrityloxy)ethyl, 1-(4,4'-dimethoxytrityloxy)-2-propyl, 3-(4,4'-dimethoxytrityloxy)-2-methyl-1-propyl and 4,4'-dimethoxytrityloxymethyl.

91. The photocleavable linker of claim 90, wherein r and s are  
5 both 0.

92. The photocleavable linker of claim 91, wherein:

R<sup>23</sup> is (diisopropylamino)(2-cyanoethoxy)P-; and

R<sup>24</sup> is selected from the group consisting of 3-(4,4'-dimethoxytrityloxy)propoxy, 4-(4,4'-dimethoxytrityloxy)butyl, 3-(4,4'-dimethoxytrityloxy)propyl, 2-(4,4'-dimethoxytrityloxy)ethyl, 1-(4,4'-dimethoxytrityloxy)-2-propyl, 3-(4,4'-dimethoxytrityloxy)-2-methyl-1-propyl and 4,4'-dimethoxytrityloxymethyl.  
10

93. The photocleavable linker of claim 92, wherein:

R<sup>24</sup> is 3-(4,4'-dimethoxytrityloxy)propoxy.

15 94. The photocleavable linker of claim 84, where in the linker is selected from the group consisting of 1-(2-nitro-5-(3-O-4,4'-dimethoxytritylpropoxy)phenyl)-1-O-((2-cyanoethoxy)-diisopropylaminophosphino)ethane and 1-(4-(3-O-4,4'-dimethoxytritylpropoxy)-3-methoxy-6-nitrophenyl)-1-O-((2-cyanoethoxy)-diisopropylaminophosphino)ethane.

1 / 123

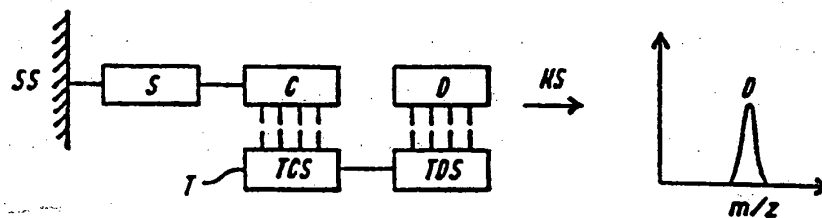


FIG. 1A

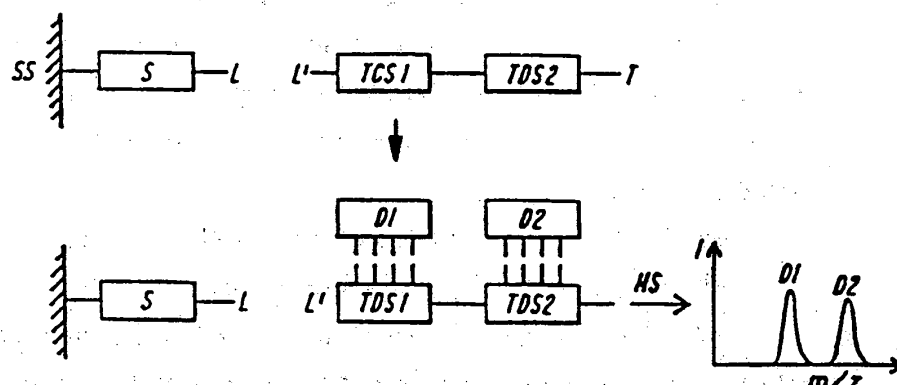


FIG. 1B

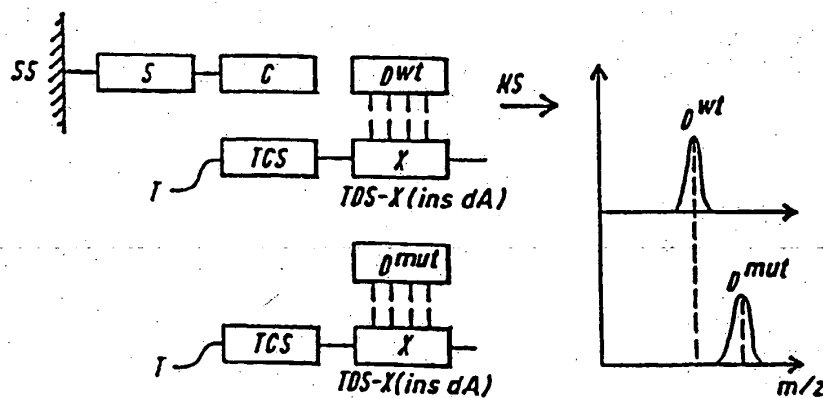
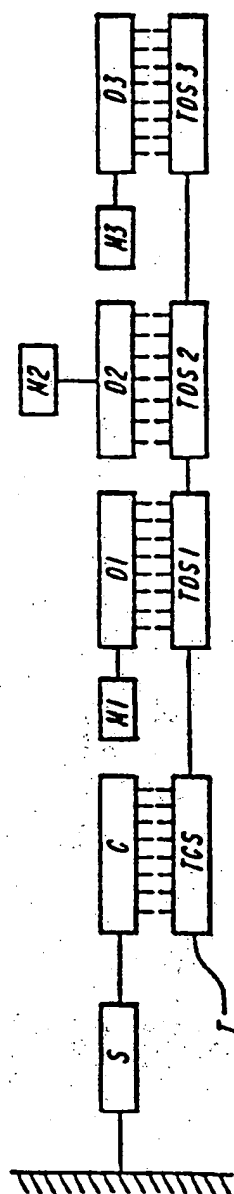


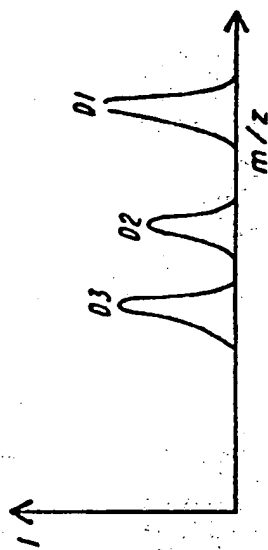
FIG. 1C



2/123



MS



$I = \text{INTENSITY}$   
 $m/z = \frac{\text{MASS}}{\text{CHARGE}}$

$H1 > H2 > H3$

$D1 - D3$  SIMILAR MOLECULAR WEIGHT

$H$  : MASS-MODIFYING FUNCTION

—: POLYMERIC SUPPORT

FIG. 2

3/123

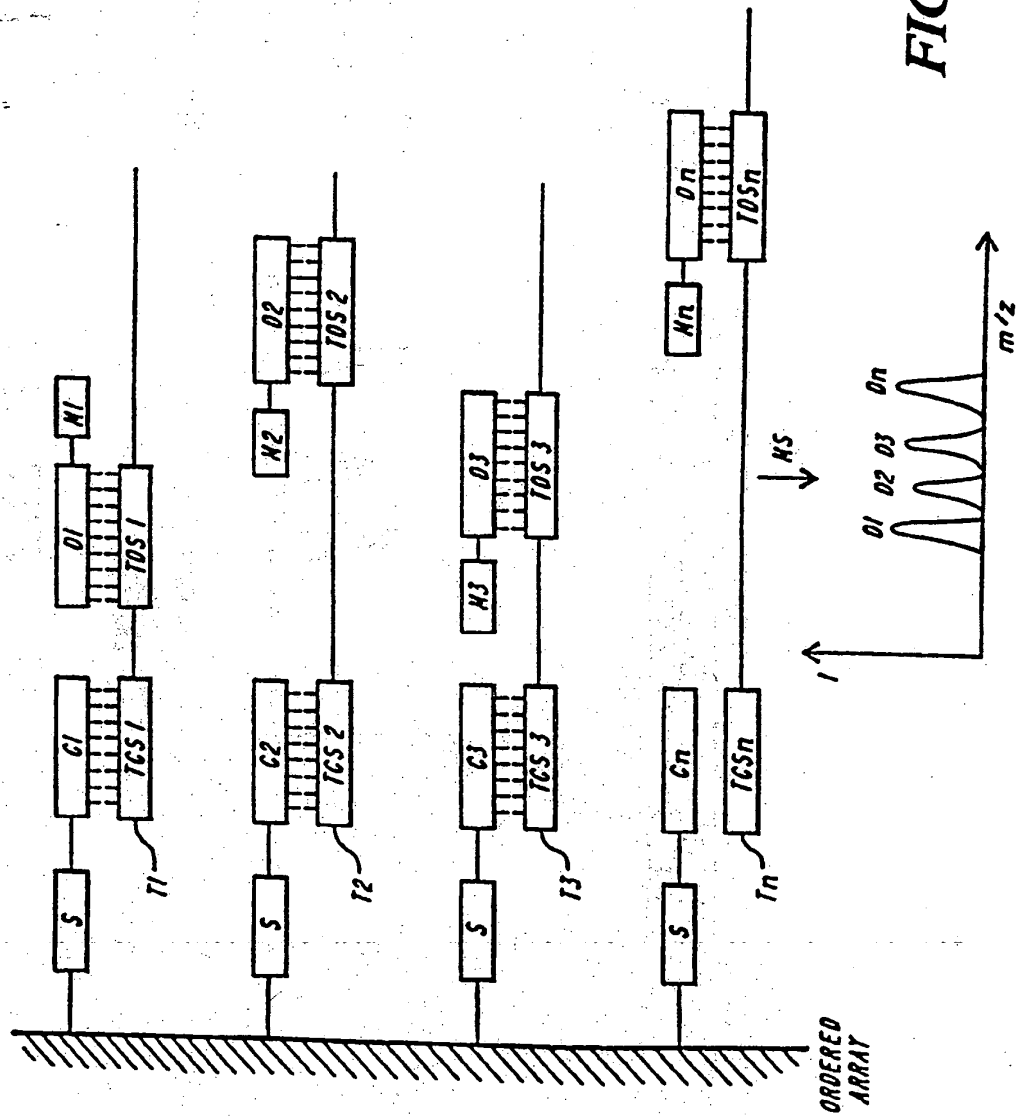


FIG. 3

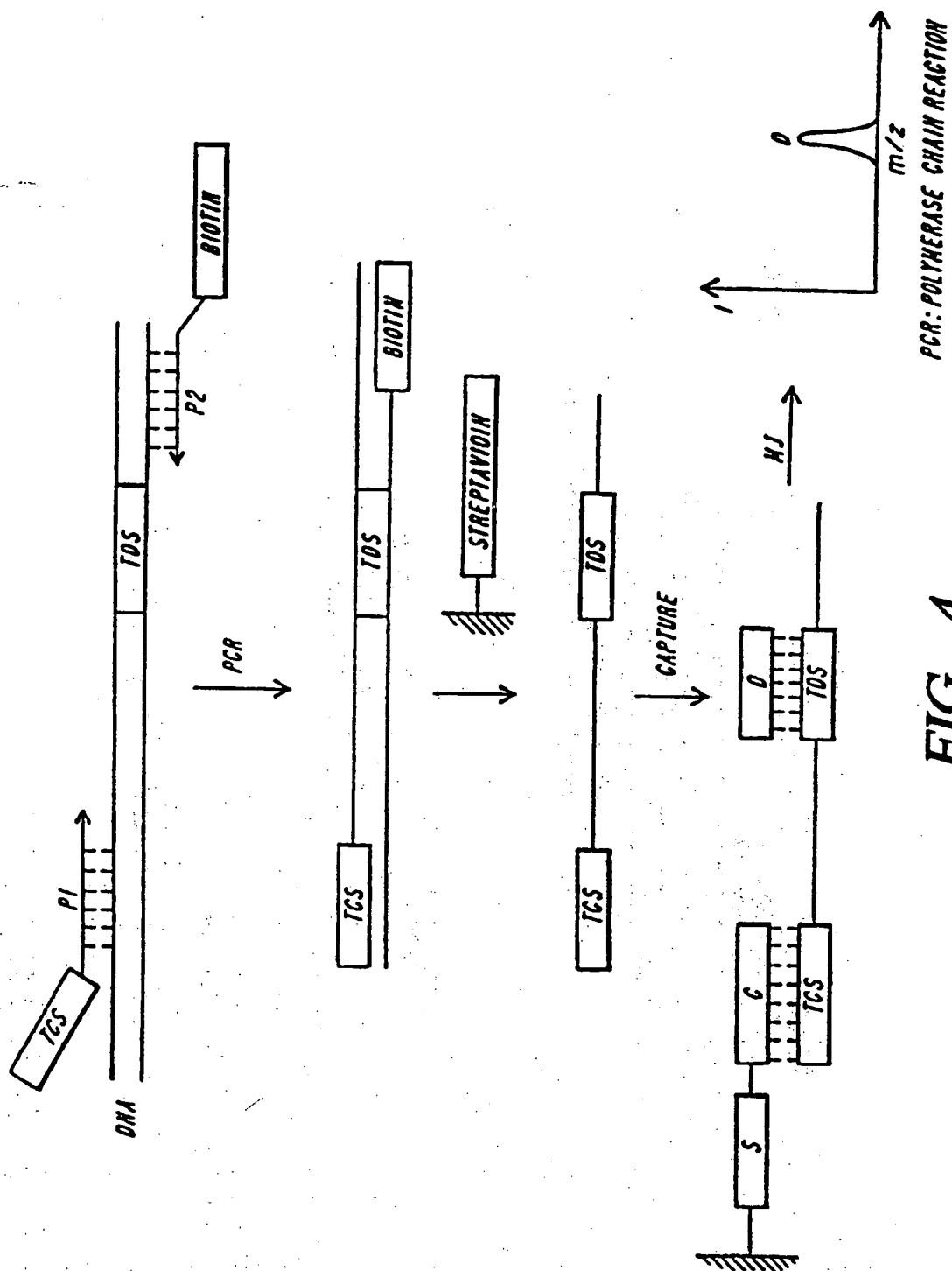
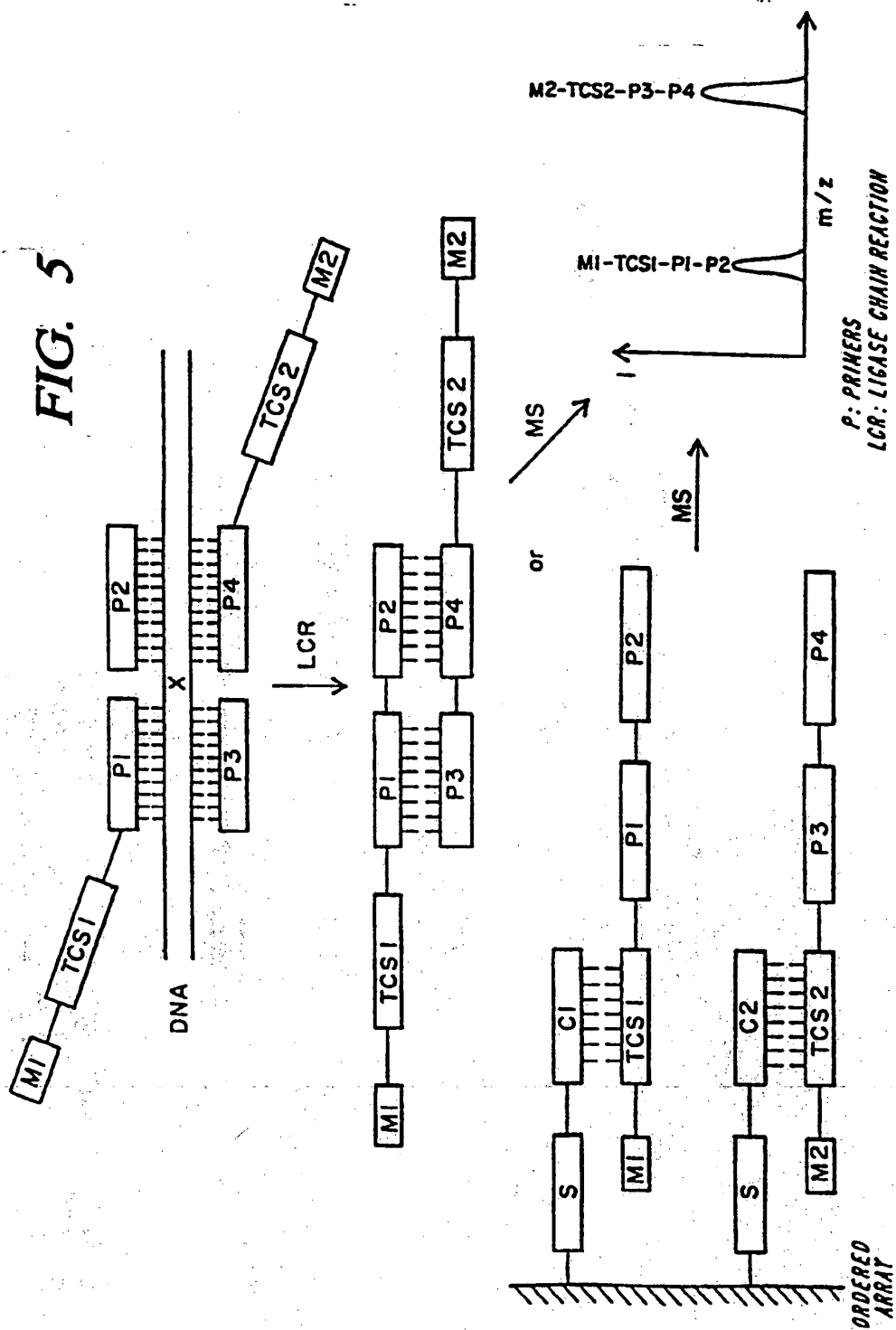


FIG. 4

FIG. 5



6/123

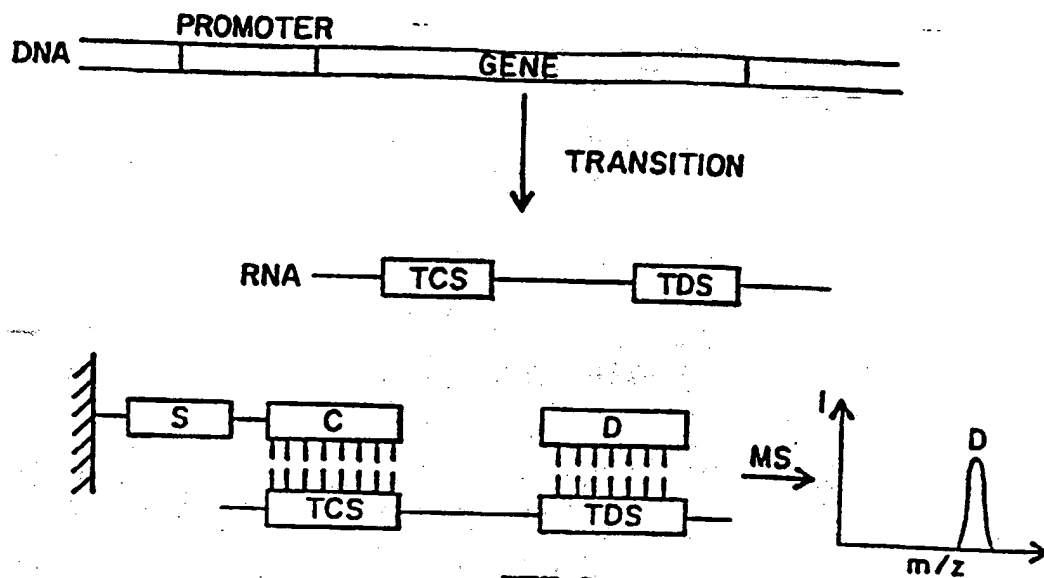


FIG. 6A

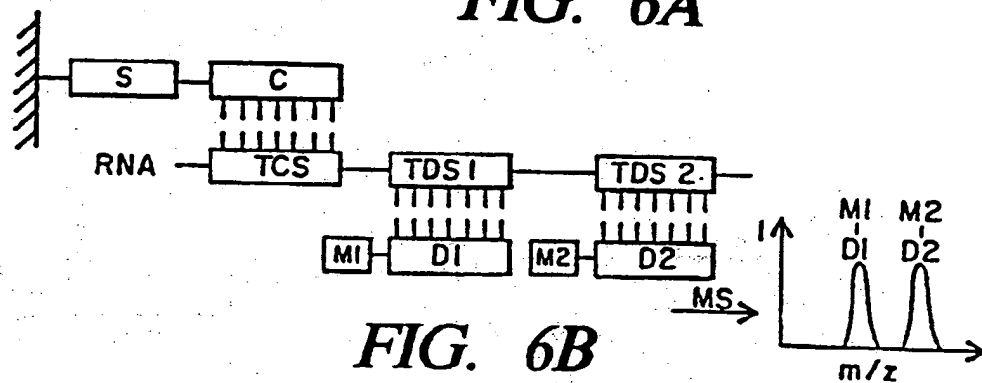


FIG. 6B

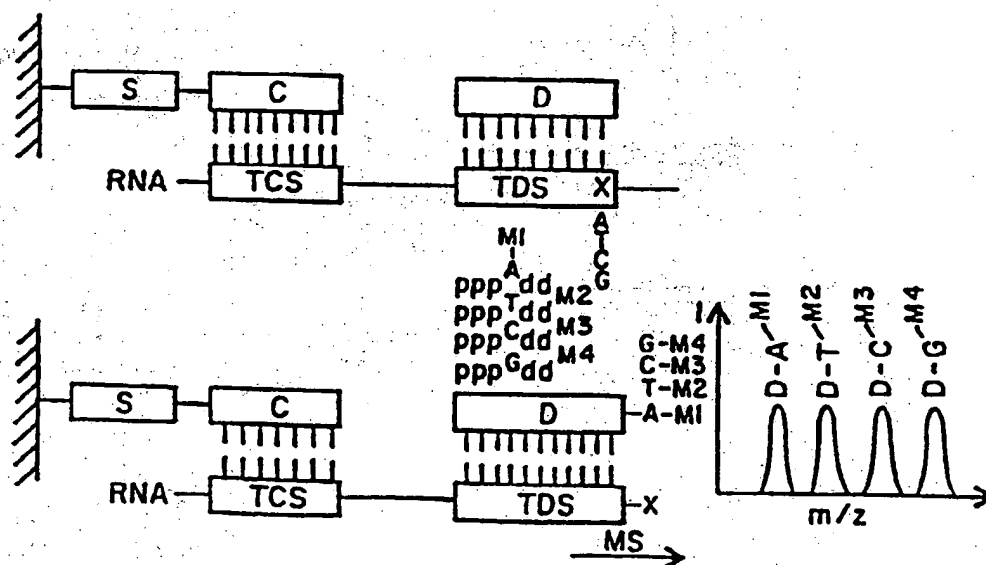
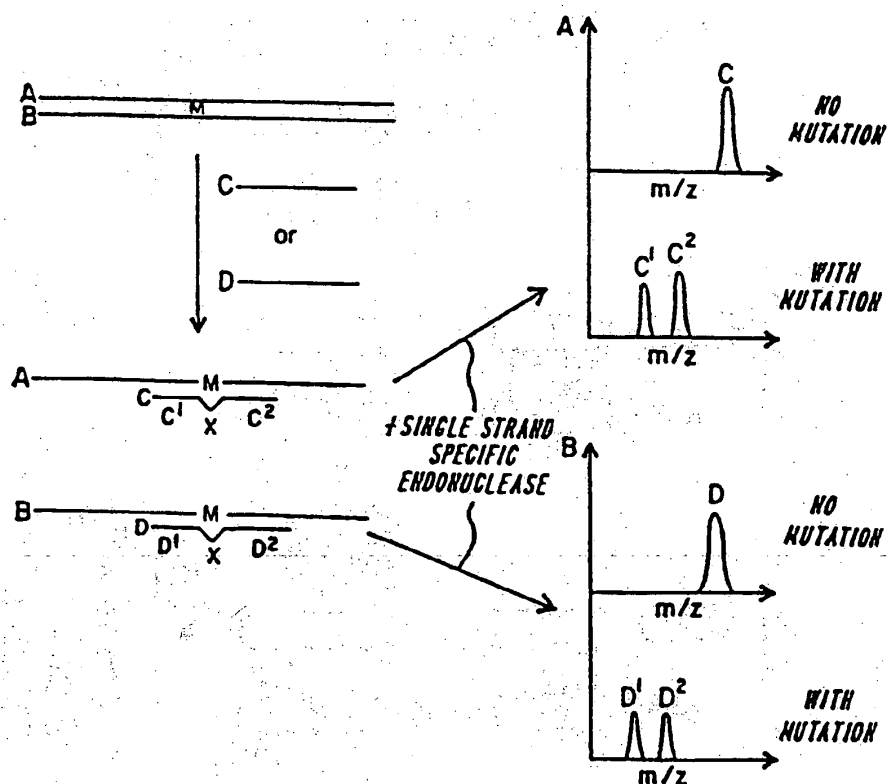
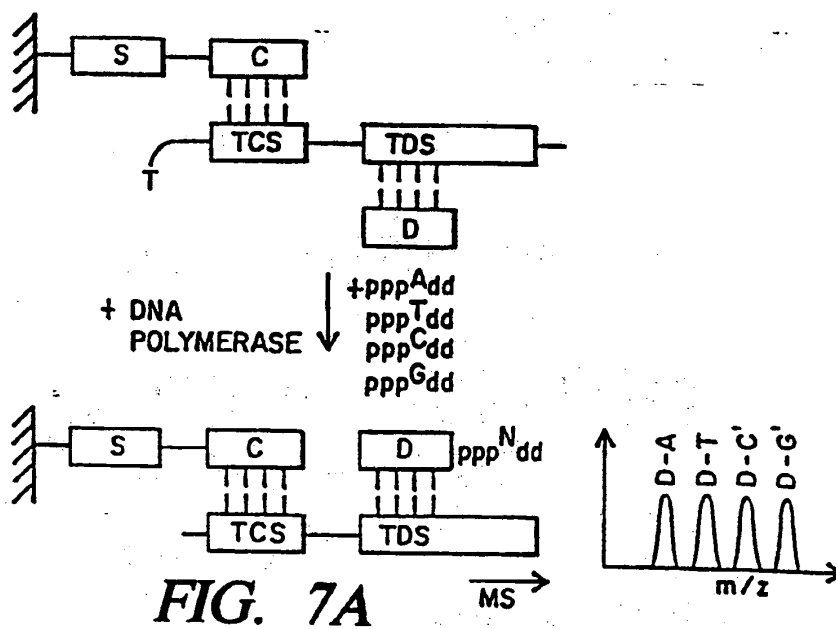
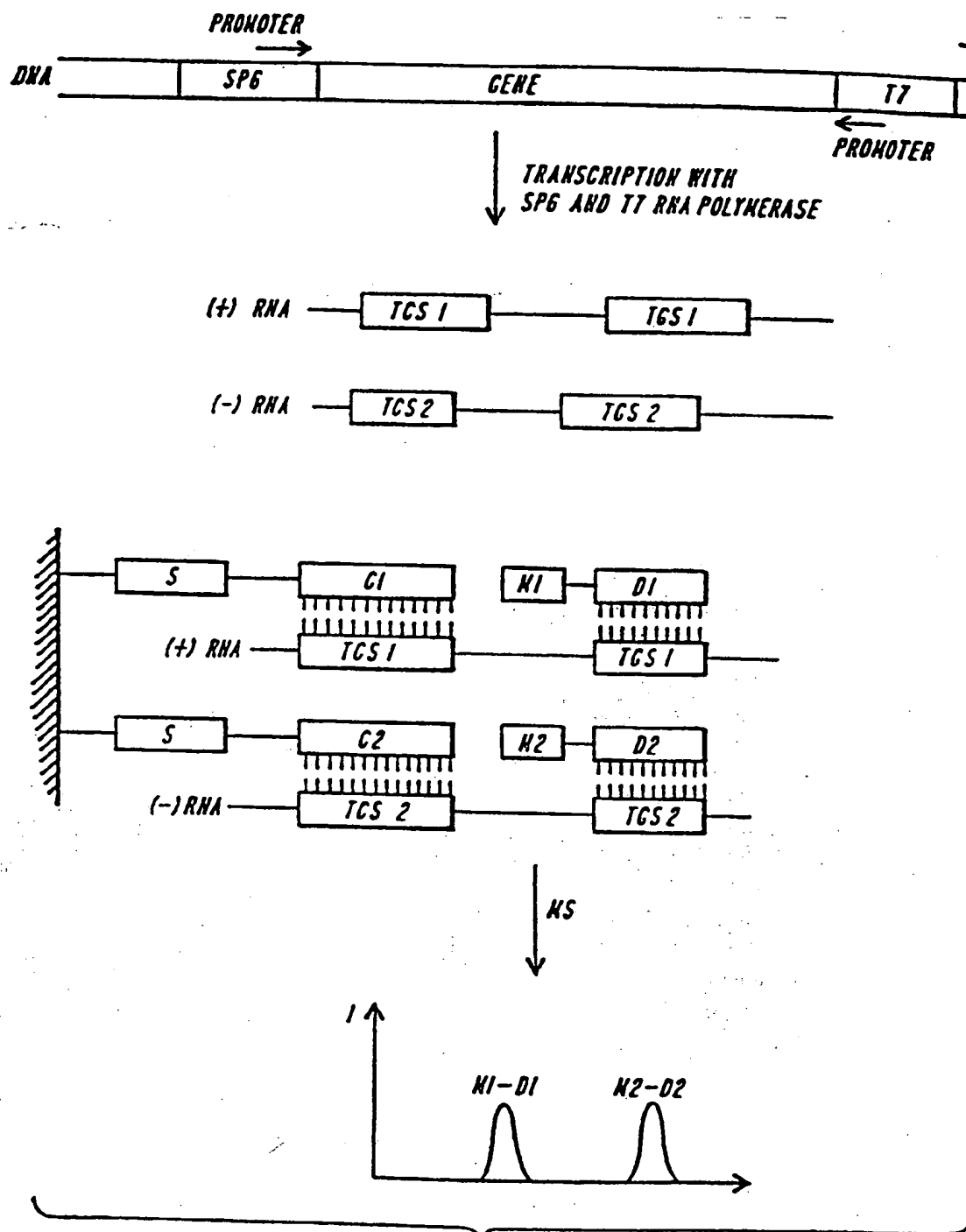


FIG. 6C

7 / 123



8/123

**FIG. 8**

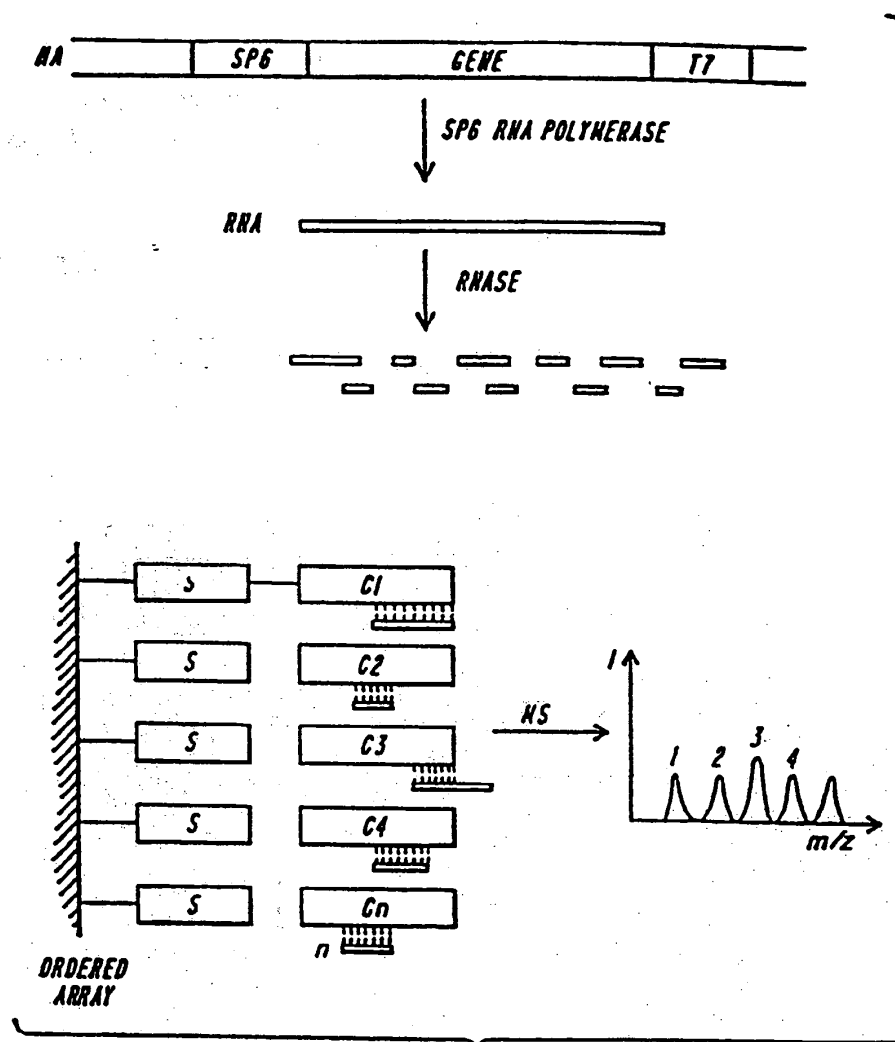


FIG. 9



10/123

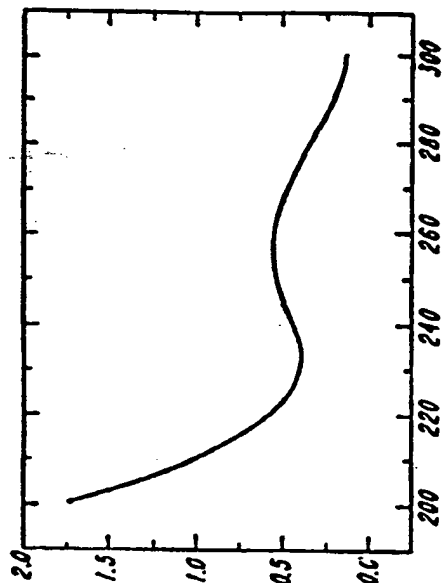


FIG. 10A-2

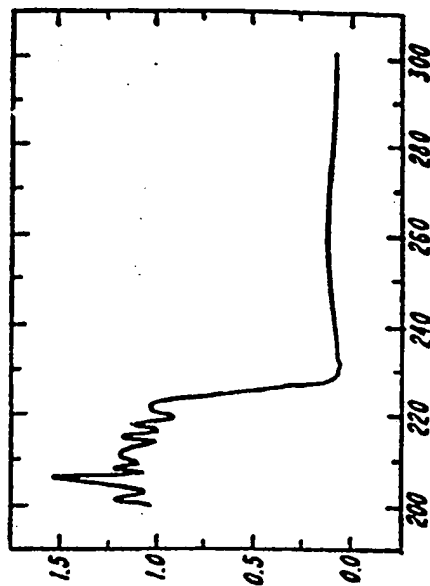


FIG. 10A-4

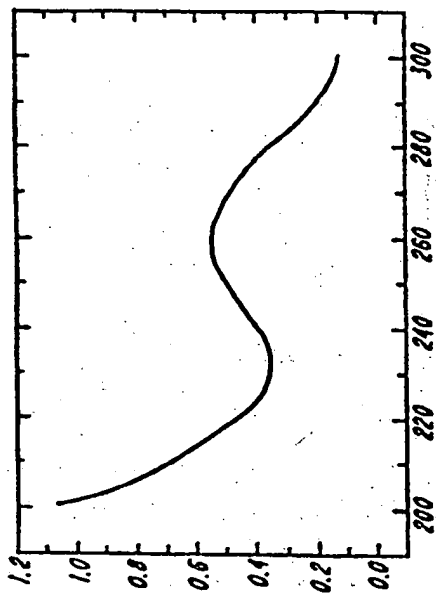


FIG. 10A-1

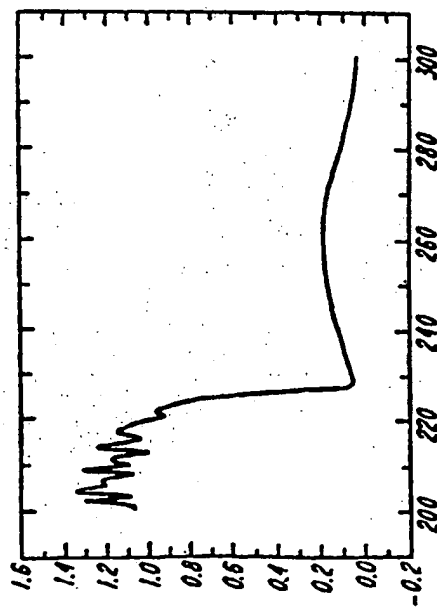
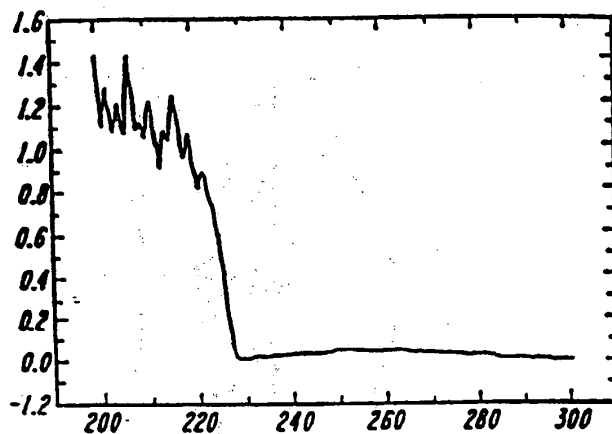
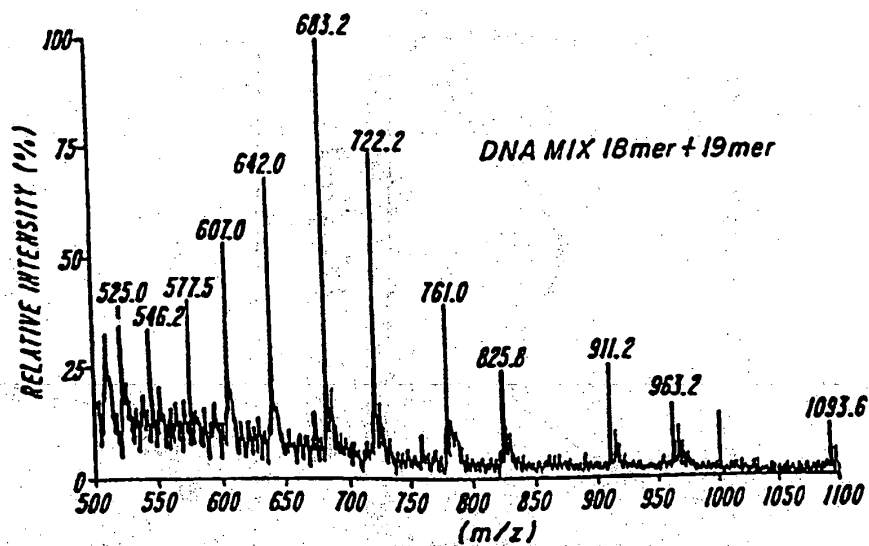


FIG. 10A-3

11 / 123

**FIG. 10B****FIG. 12A**

12/123

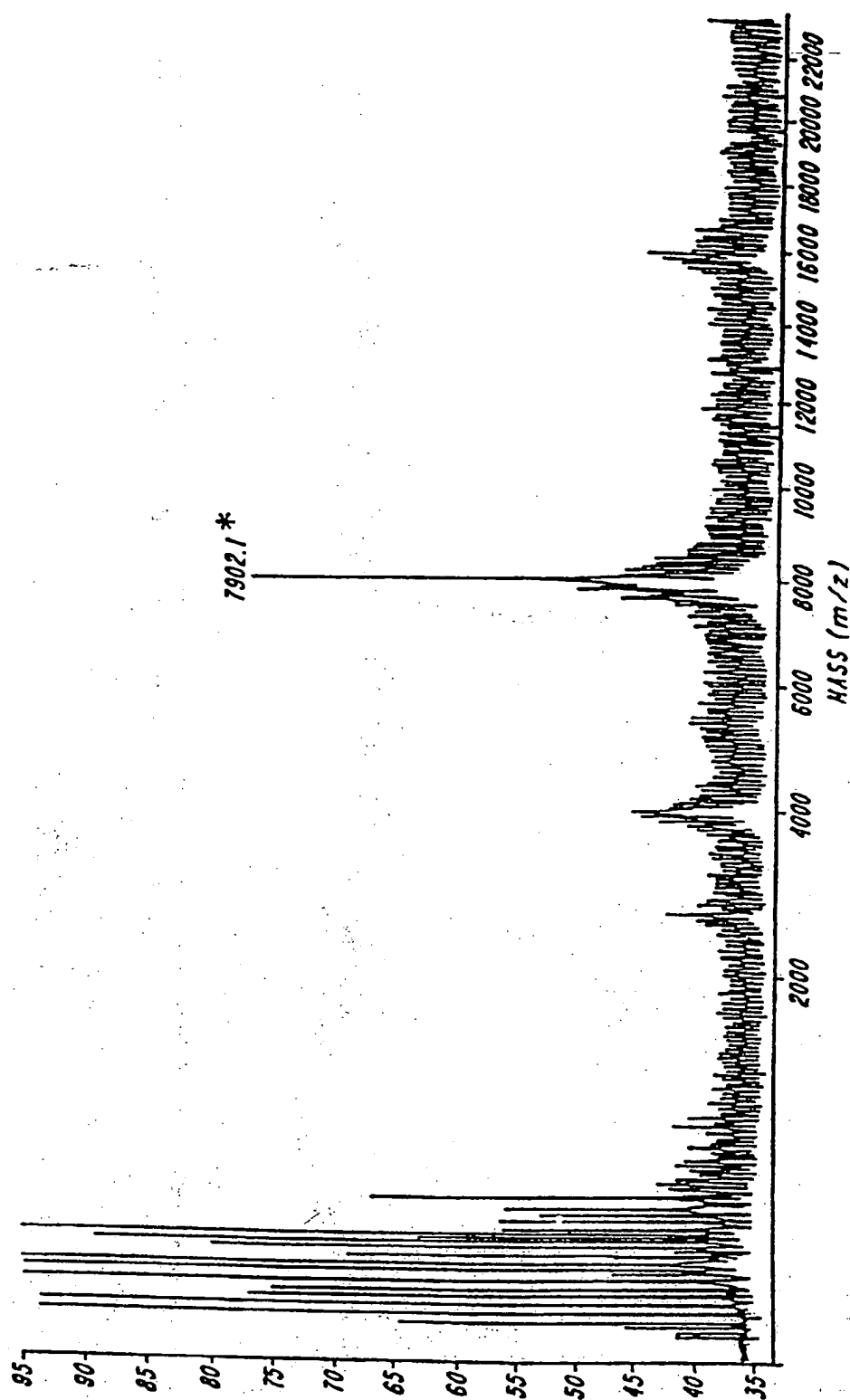


FIG. 10C

13 / 123

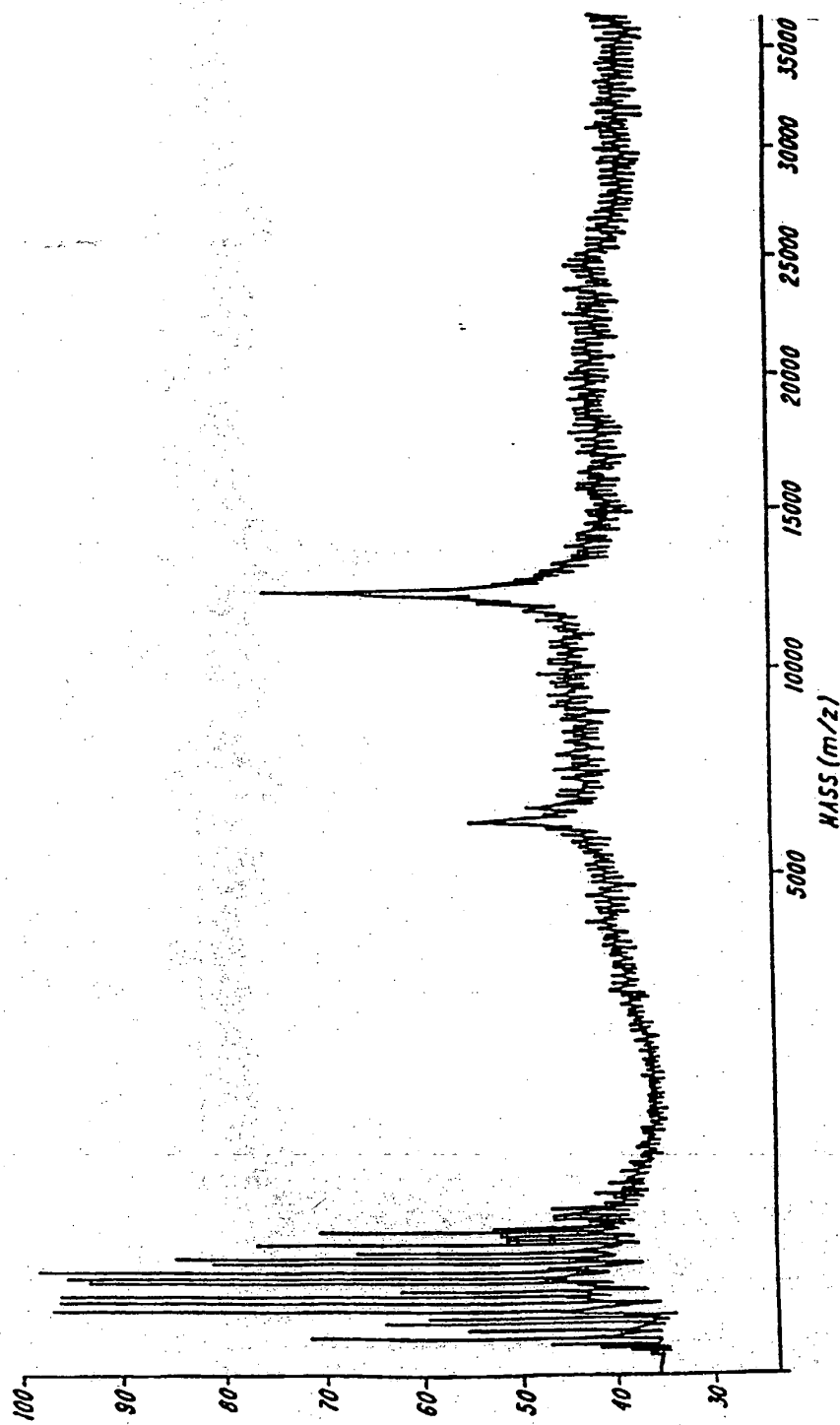
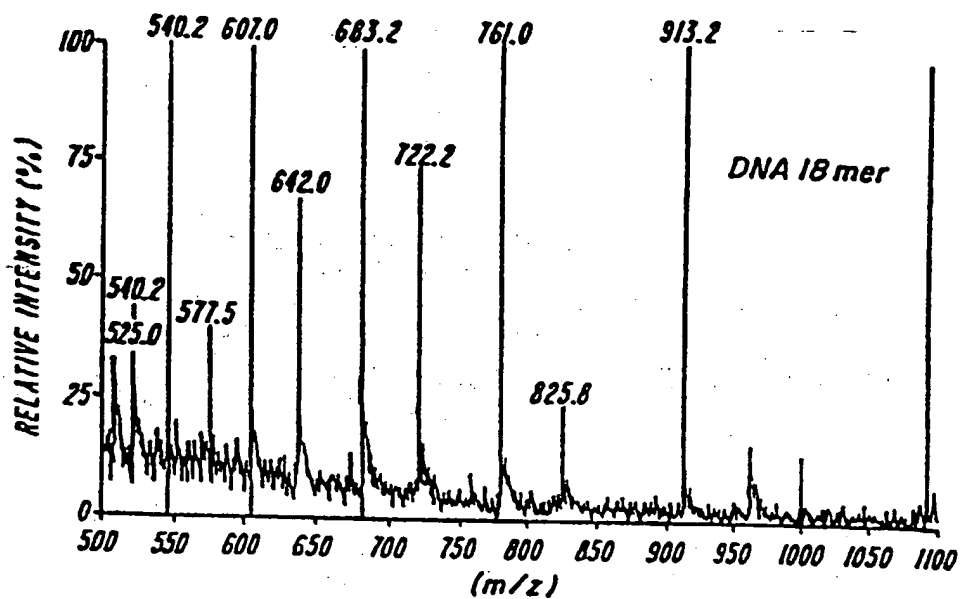
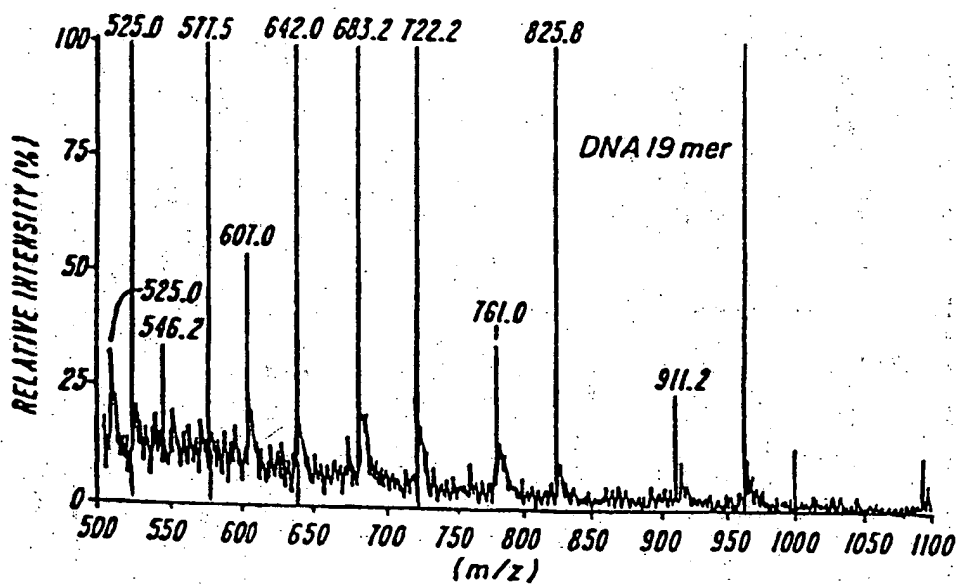


FIG. 11

SUBSTITUTE SHEET (RULE 26)

14 / 123

**FIG. 12B****FIG. 12C**

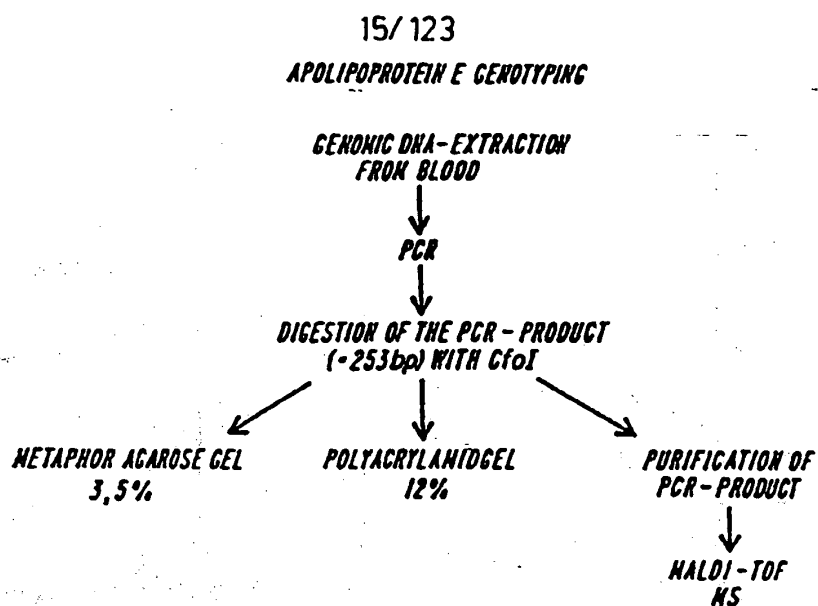


FIG. 19

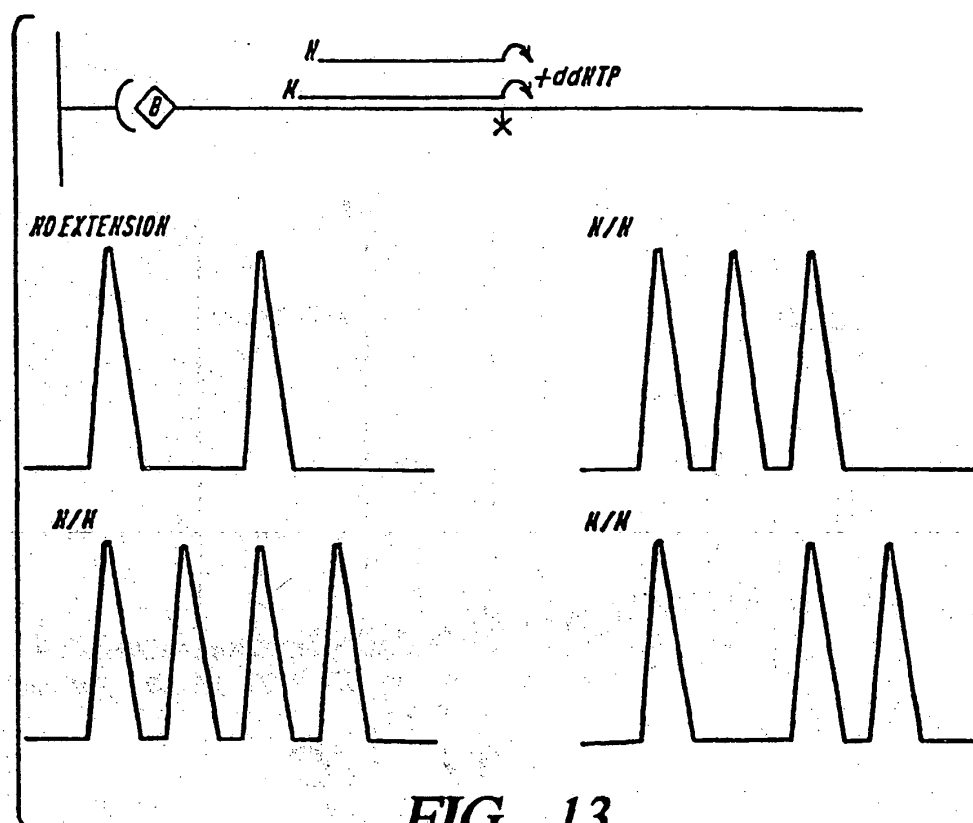


FIG. 13

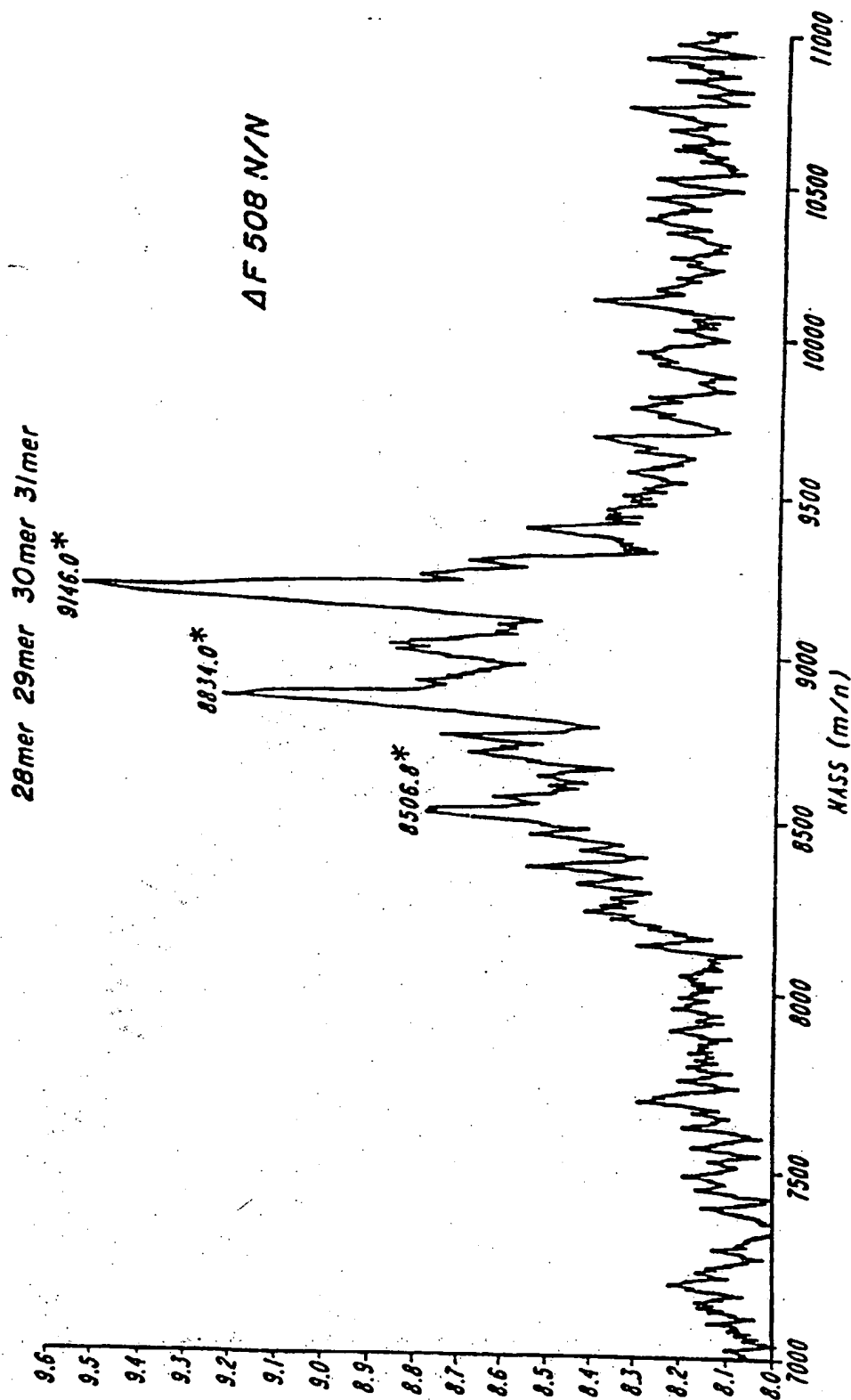


FIG. 14

17/123

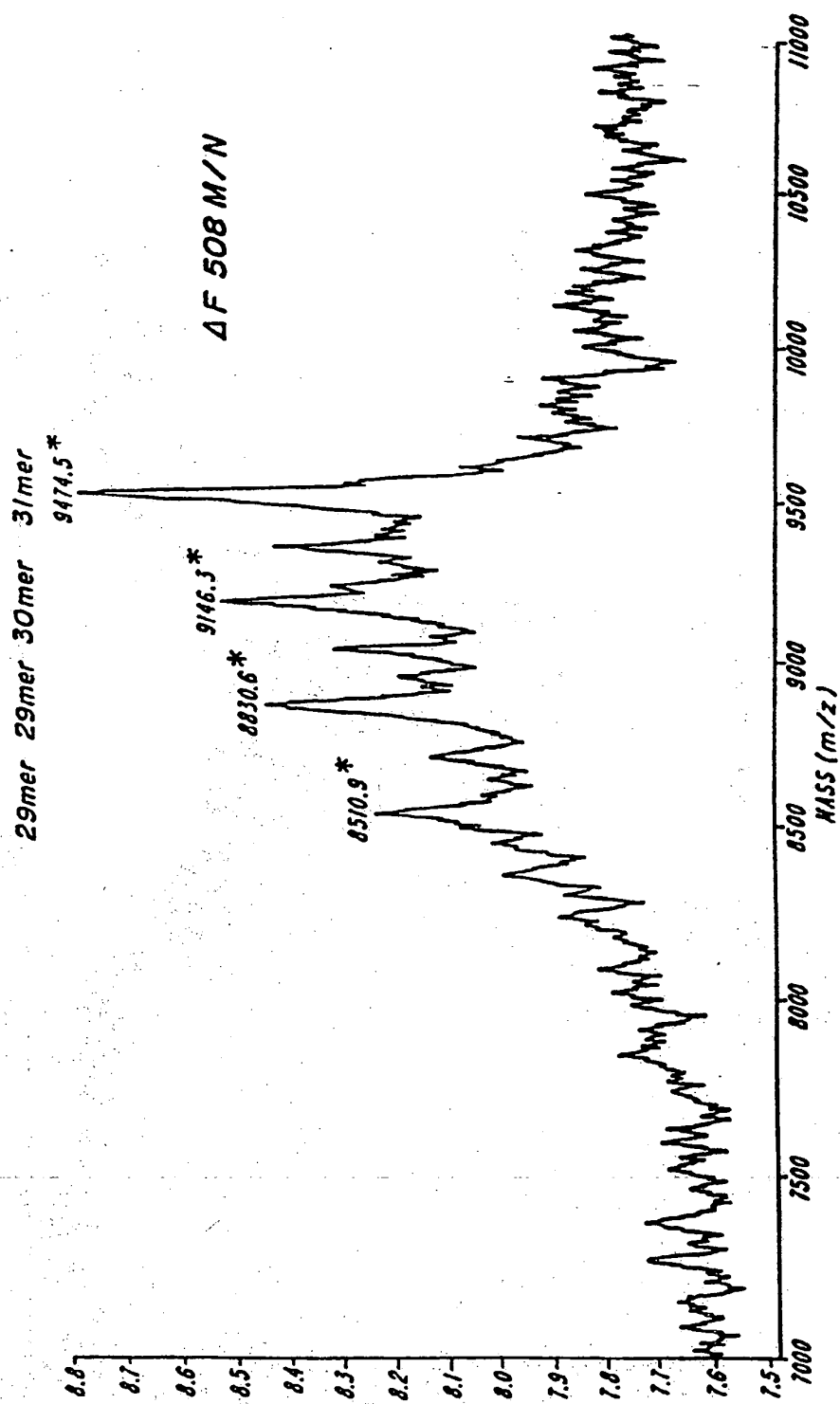


FIG. 15



18 / 123

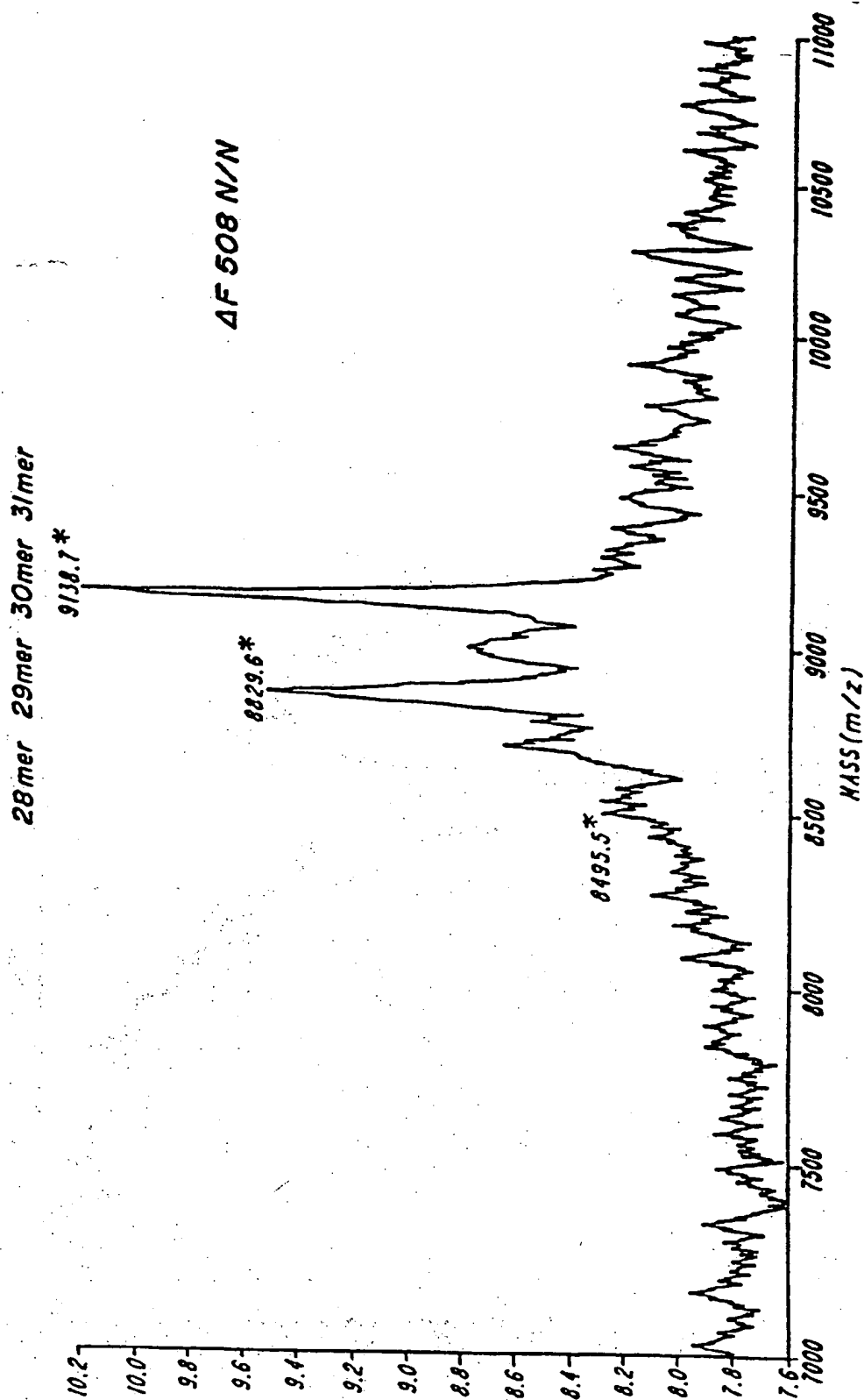


FIG. 16

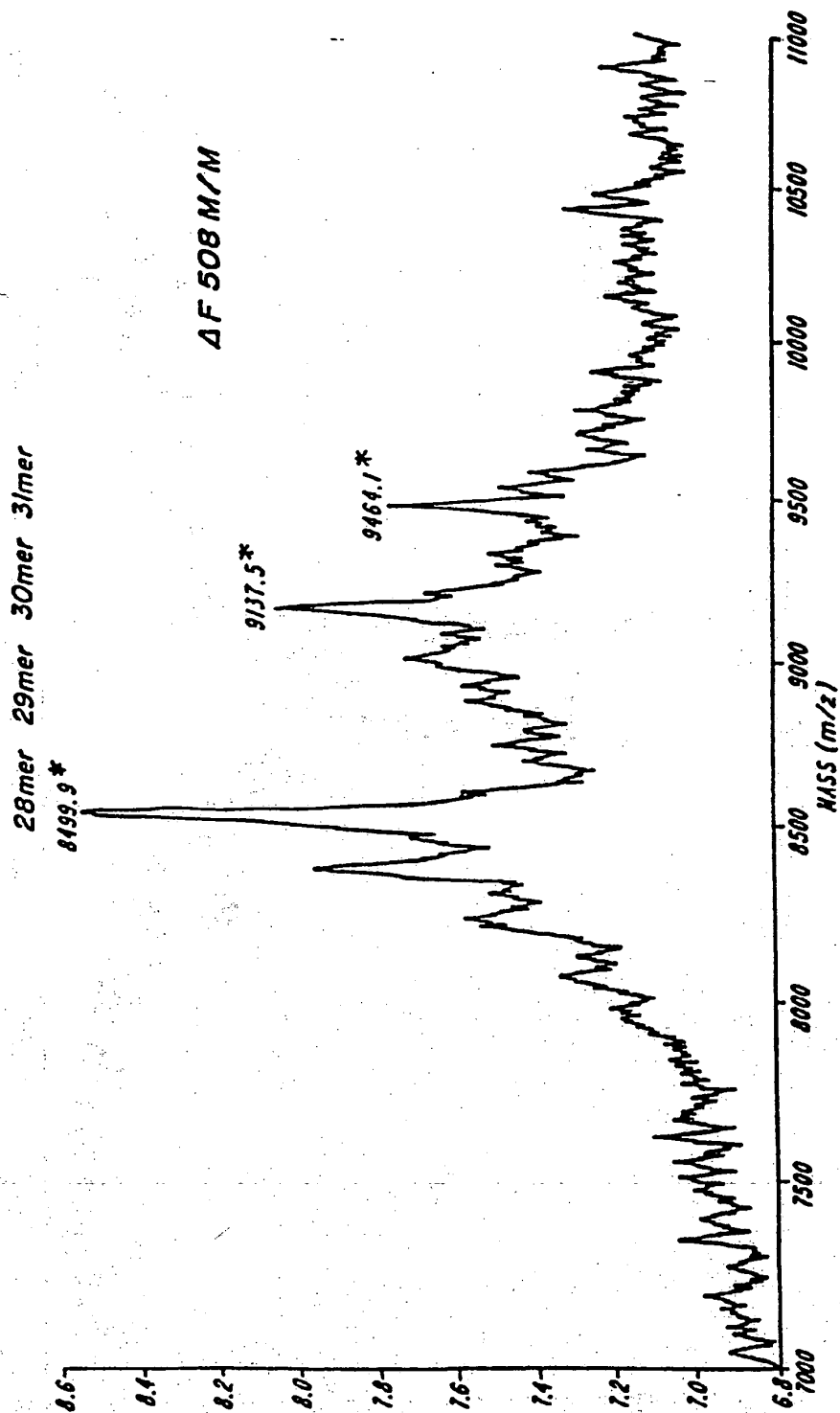


FIG. 17

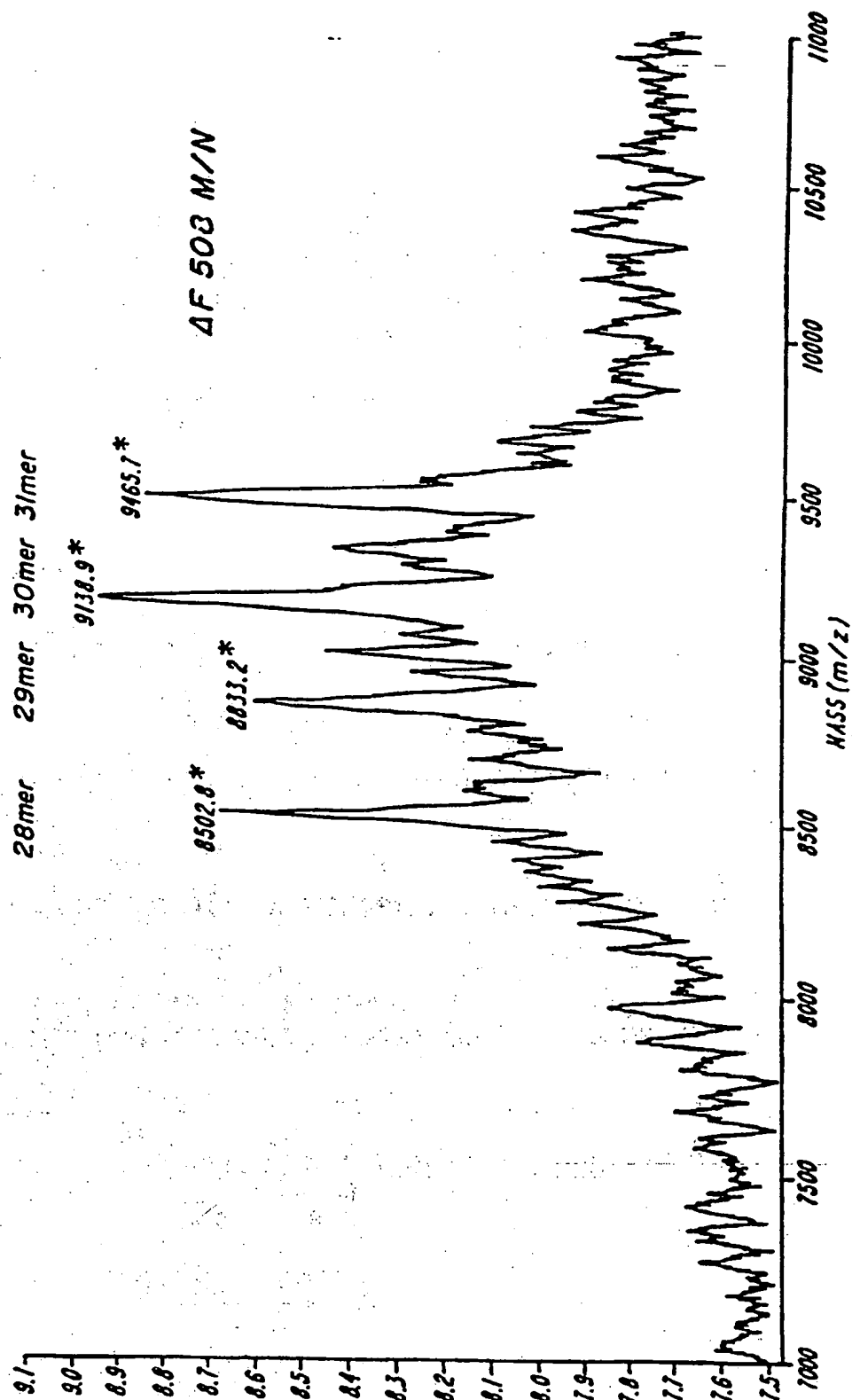


FIG. 18

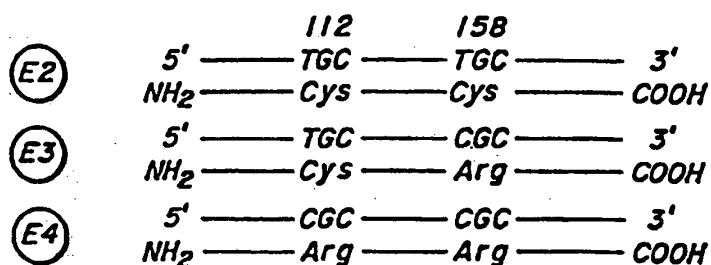


FIG. 20A

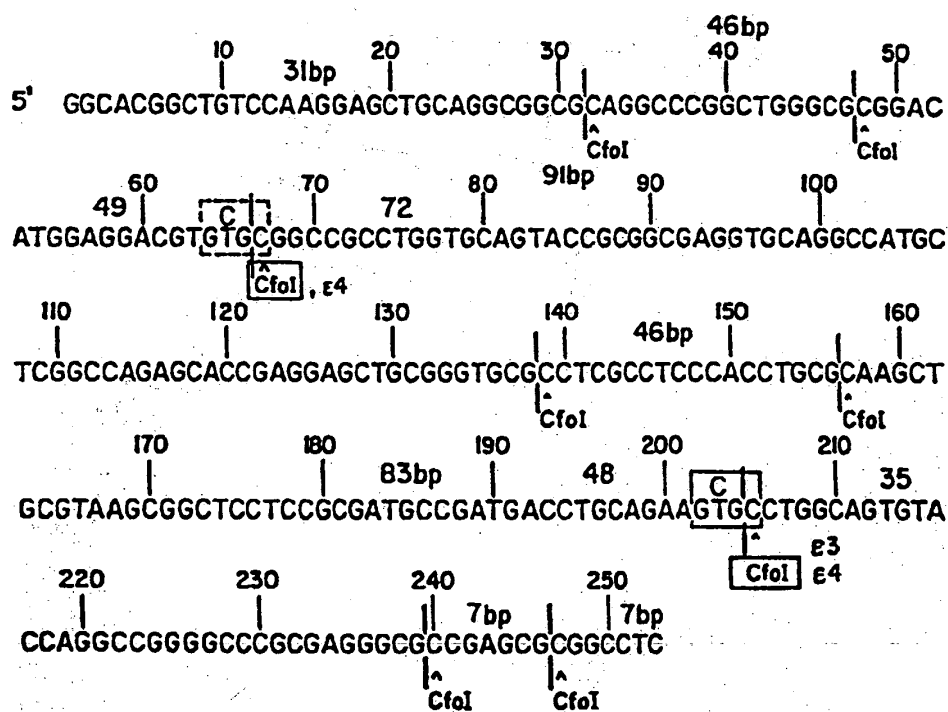
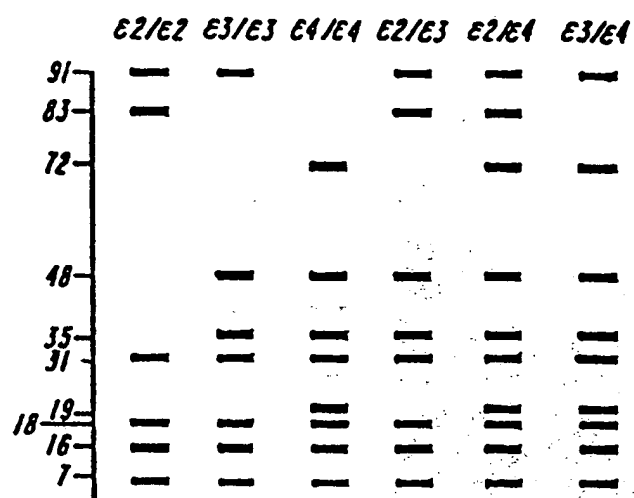
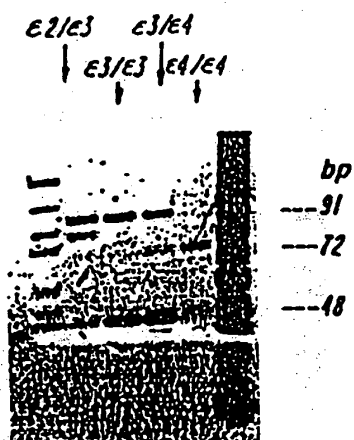


FIG. 20B

22/123

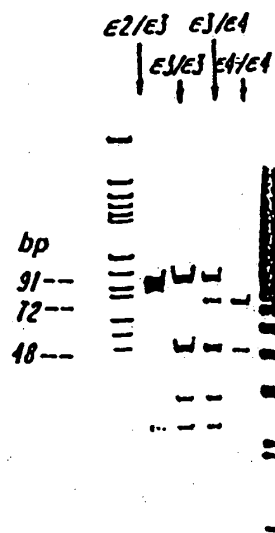
**FIG. 21A**

*MetaPhor Agarose Gel*  
3, 5 %



**FIG. 21B**

*Polyacrylamidgel*  
12 %



**FIG. 21C**

24/123

MOLECULAR WEIGHT OF THE VARIABLE FRAGMENTS IN Da:

			$\epsilon 2/\epsilon 2$	$\epsilon 3/\epsilon 3$	$\epsilon 4/\epsilon 4$	$\epsilon 2/\epsilon 3$	$\epsilon 2/\epsilon 4$	$\epsilon 3/\epsilon 4$
91 bp	SENSE	28421	X	X		X	X	X
	ANTISENSE	27864						
83bp	SENSE	25747	X			X	X	
	ANTISENSE	25591						
72bp	SENSE	22440			X		X	X
	ANTISENSE	21494						
48bp	SENSE	14844		X	X	X	X	X
	ANTISENSE	14857						
35bp	SENSE	10921		X	X	X	X	X
	ANTISENSE	10751						

FIG. 22A

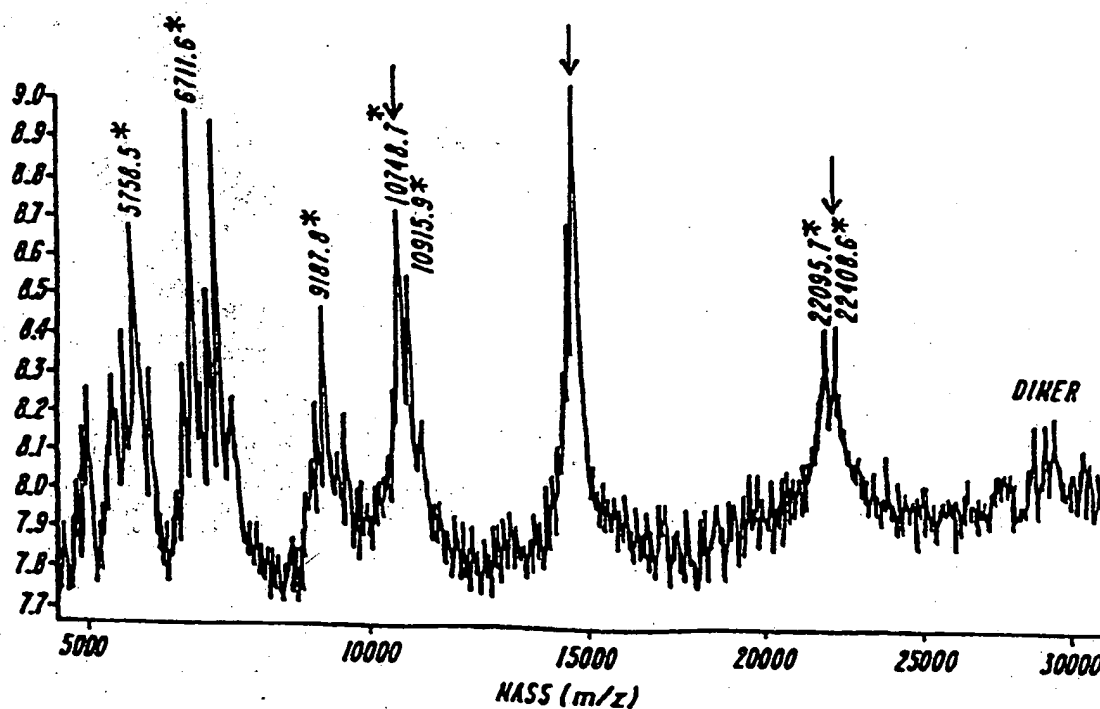


FIG. 22B

25/123

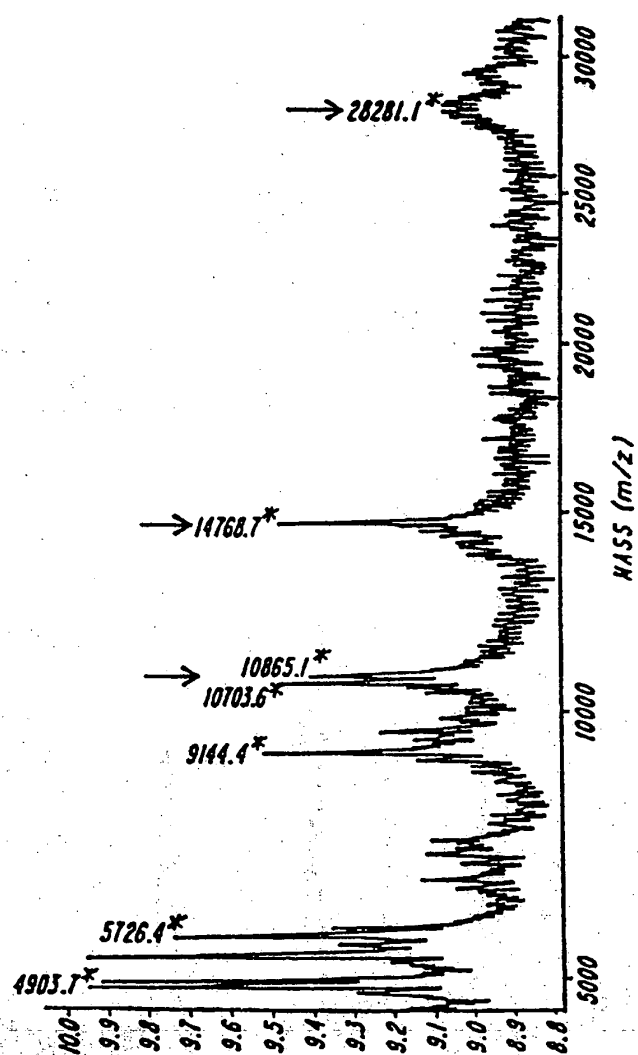


FIG. 23A



26 / 123

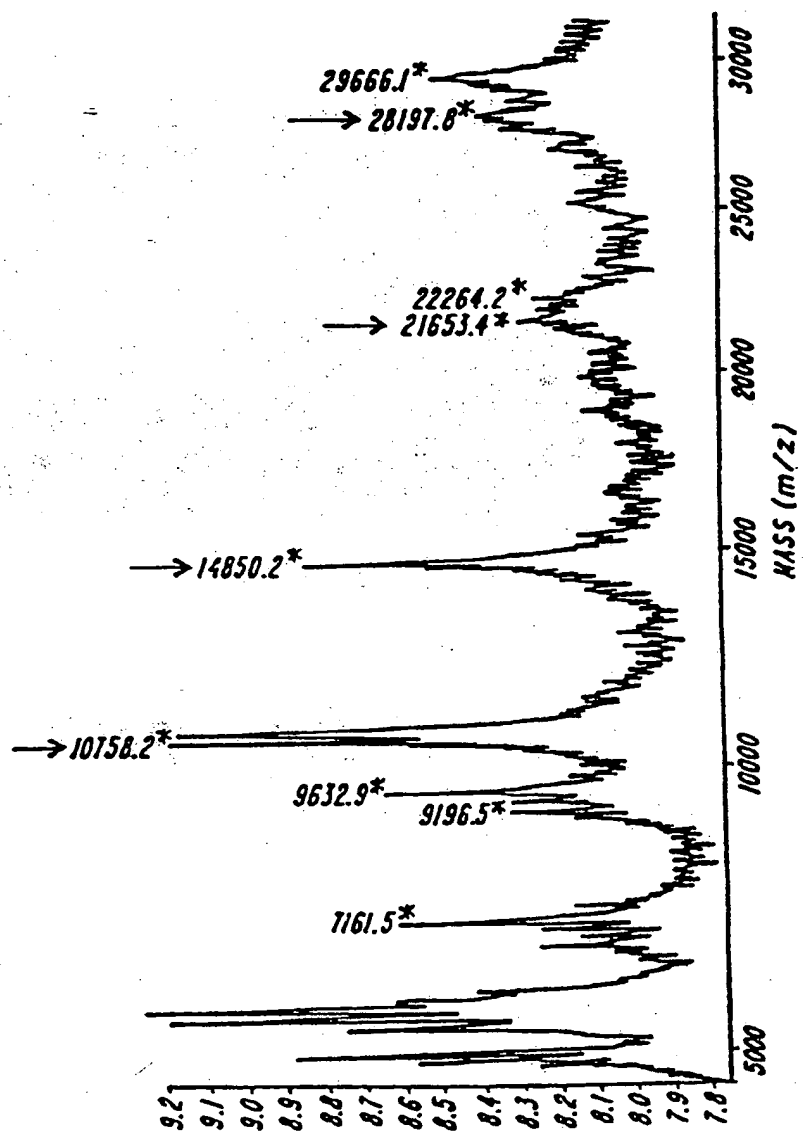
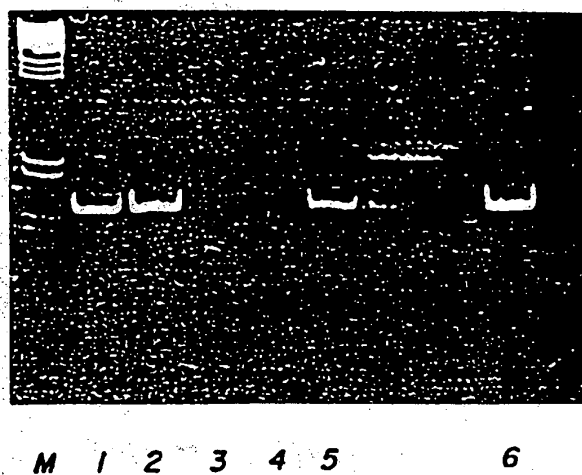
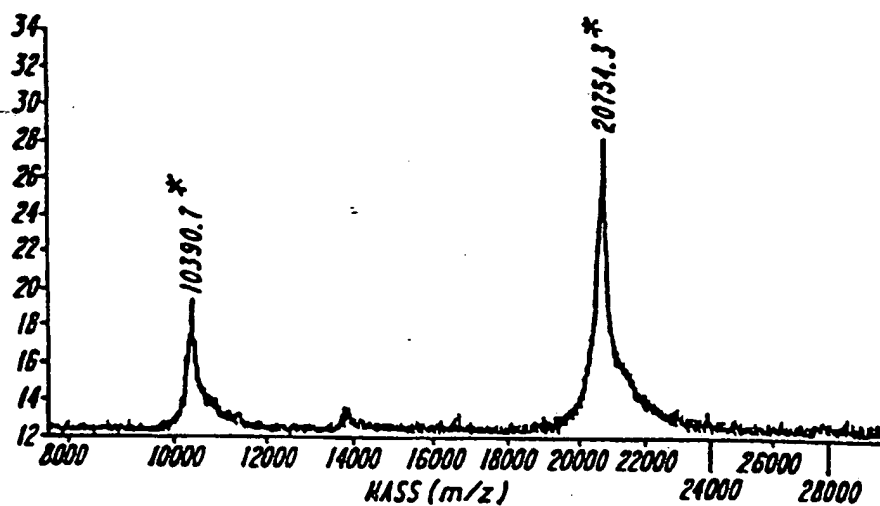
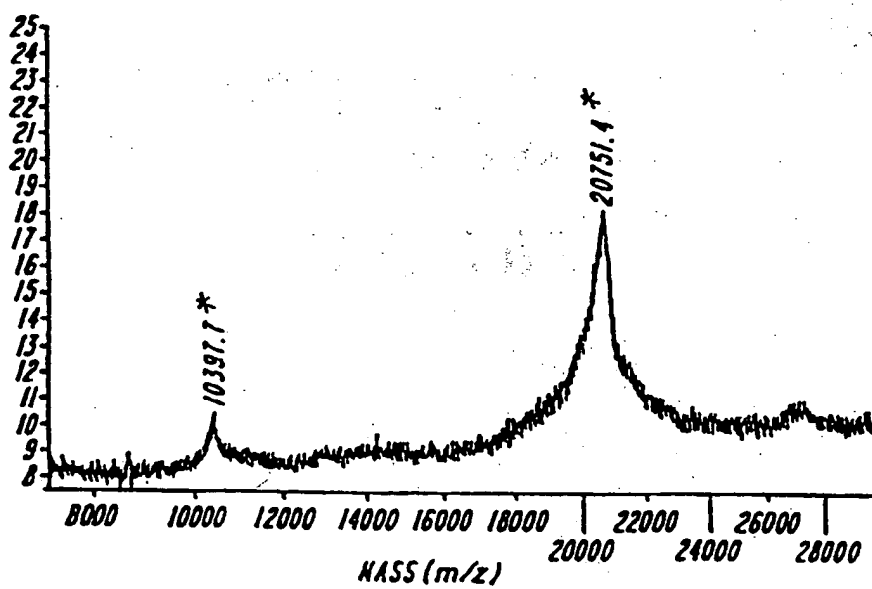


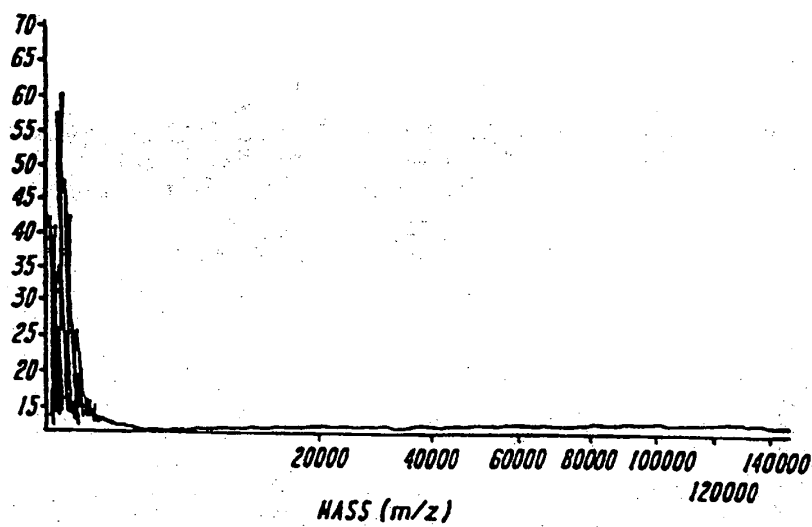
FIG. 23B

**FIG. 24**

28 / 123

*FIG. 25A**FIG. 25B*

29 / 123

*FIG. 25C*

30/123

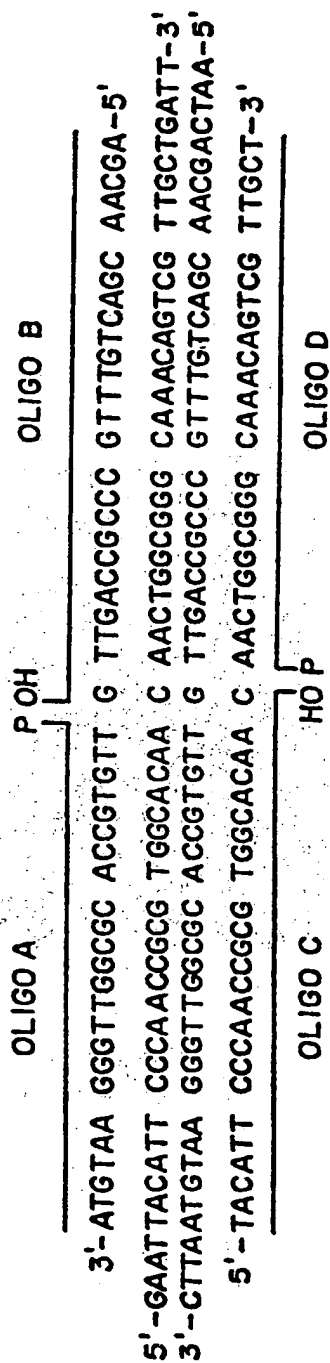
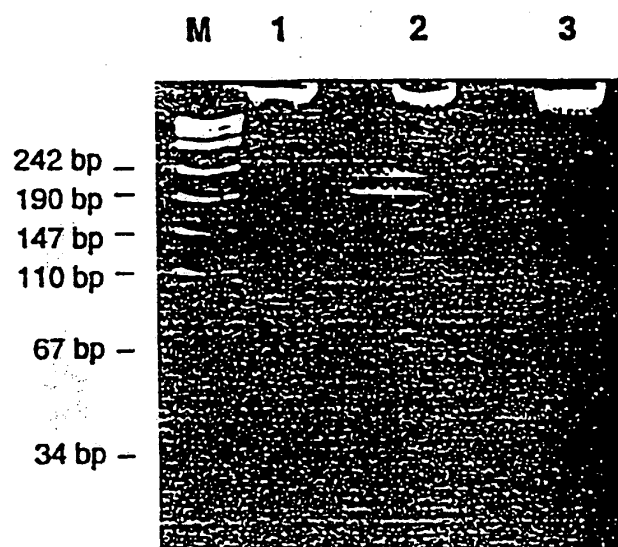
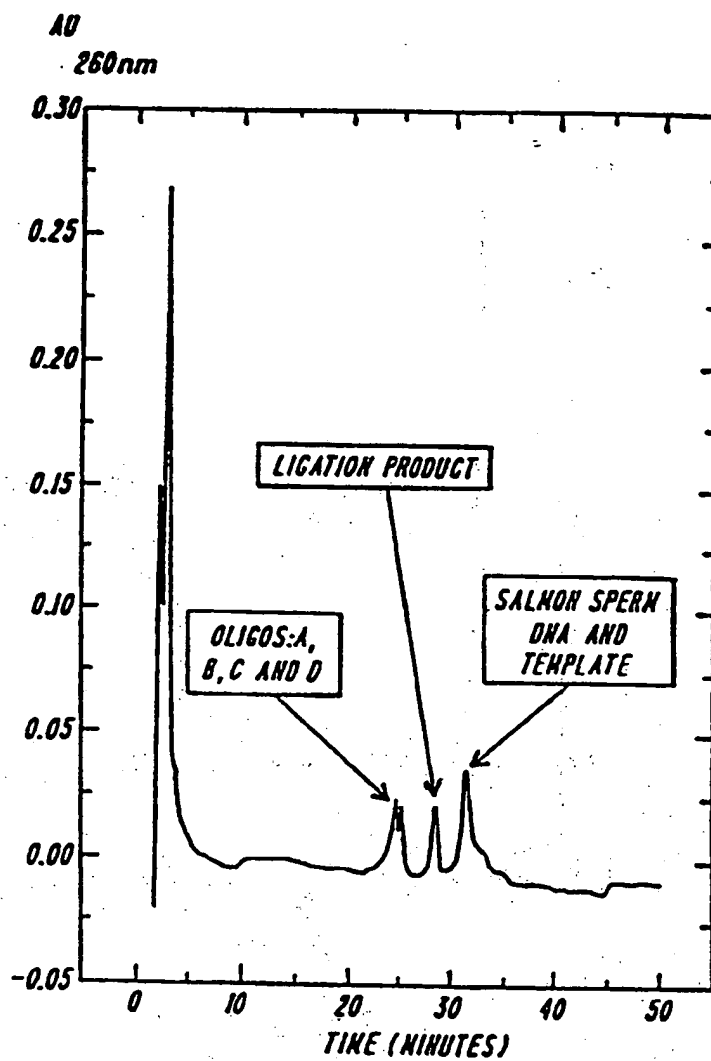


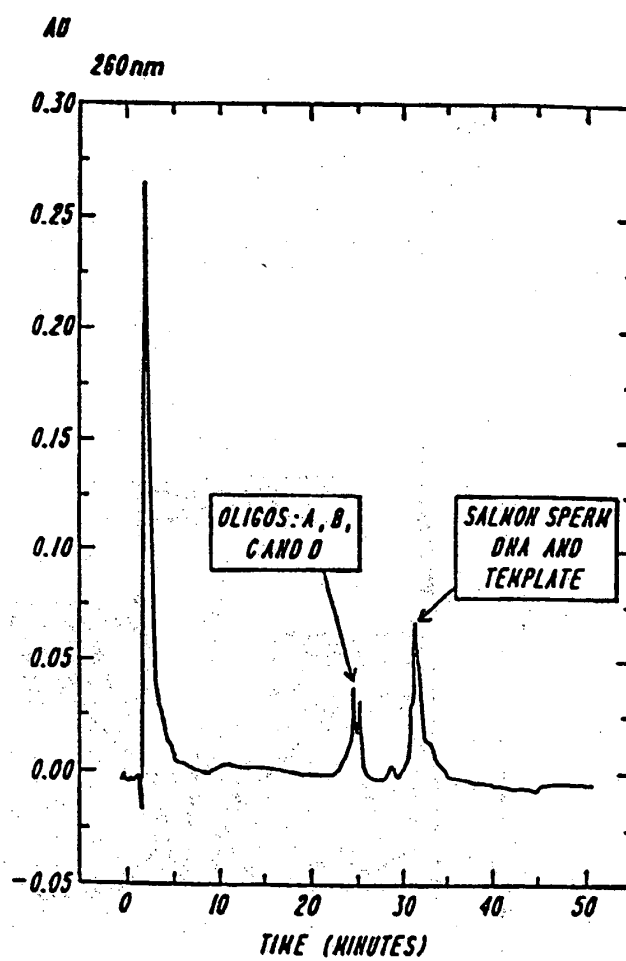
FIG. 26

31 / 123

**FIG. 27**

32/123

**FIG. 28**

**FIG. 29**



34 / 123

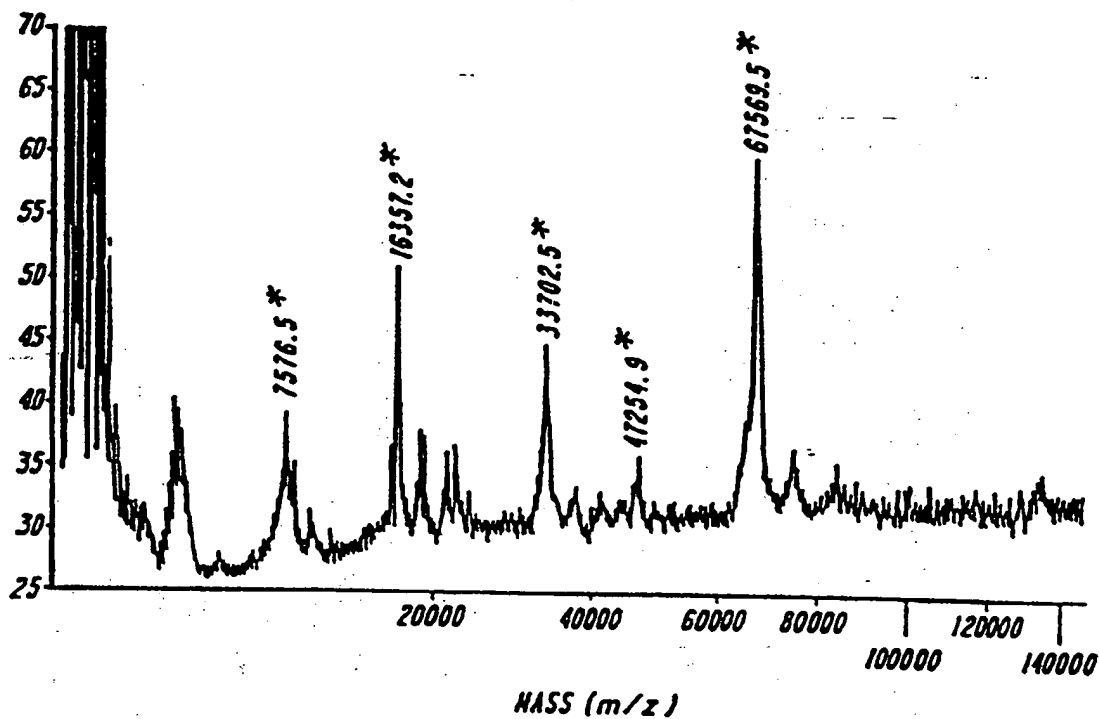


FIG. 30A

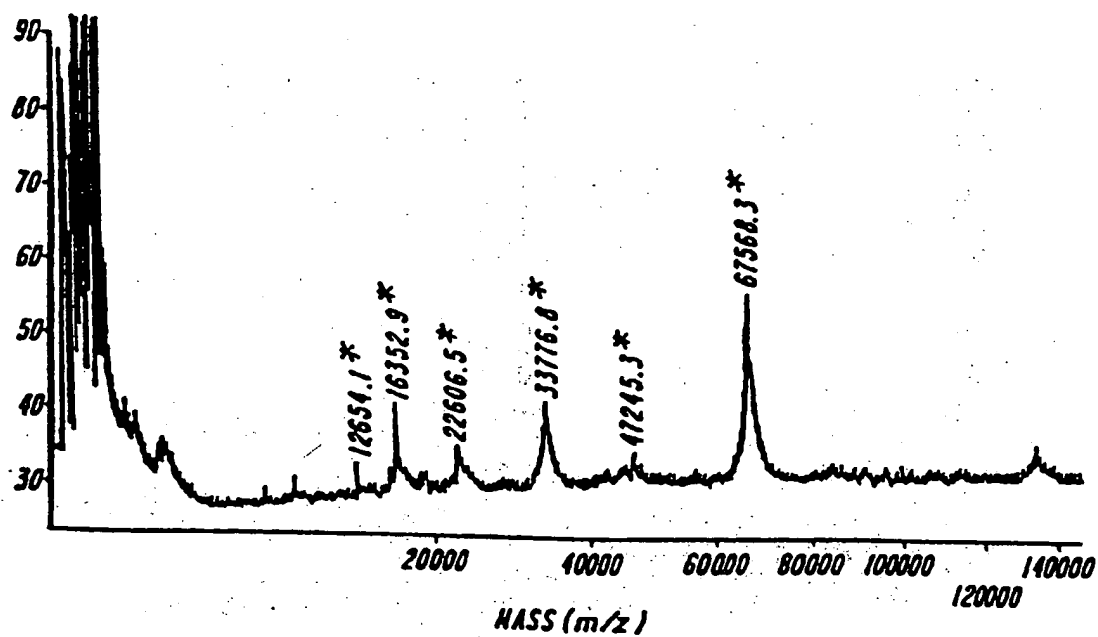


FIG. 30B

35/123

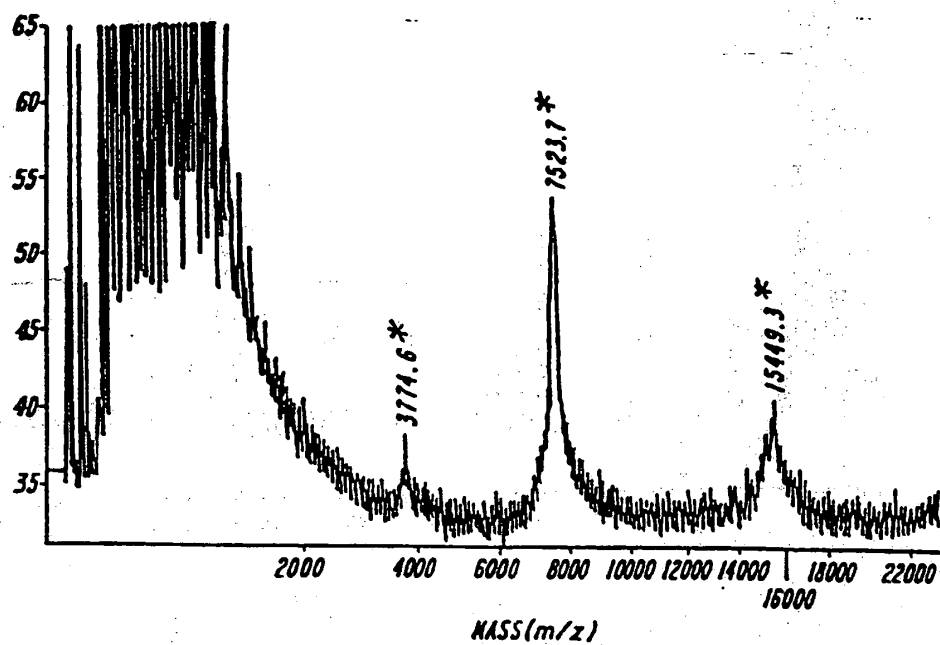


FIG. 31A

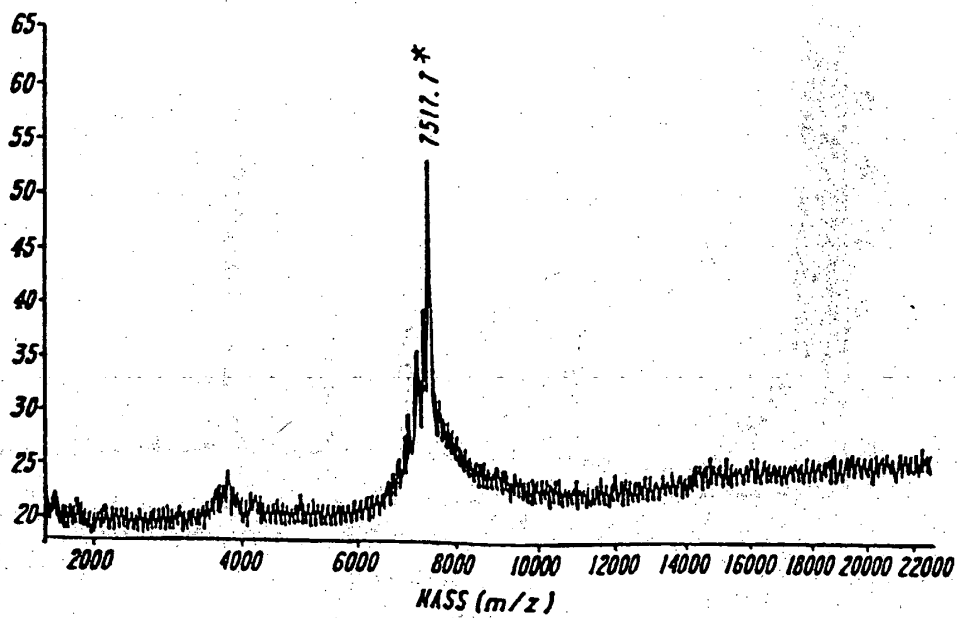
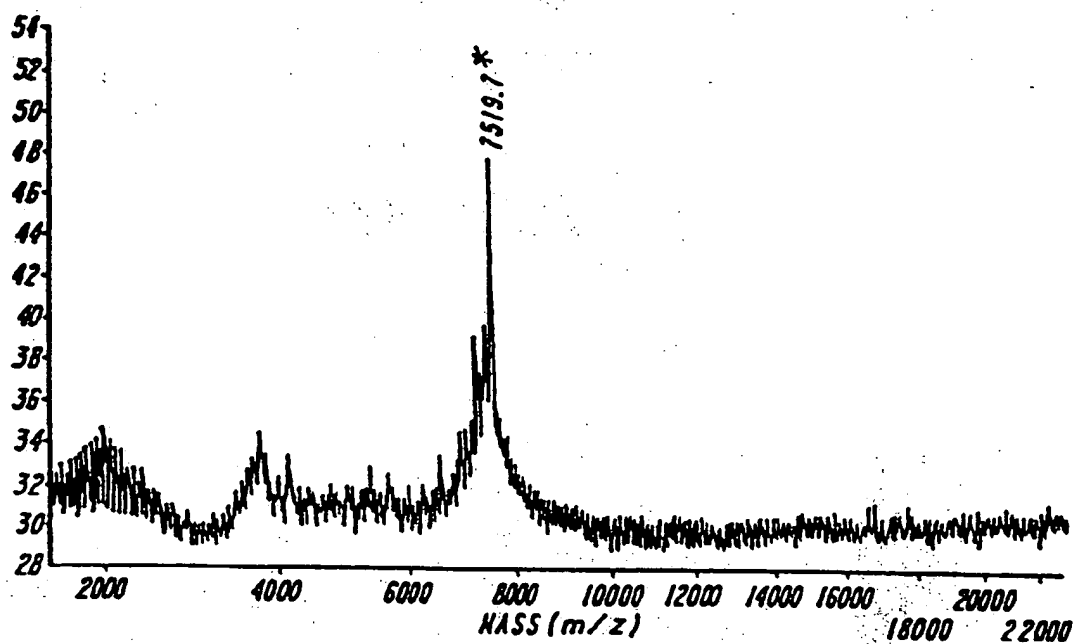
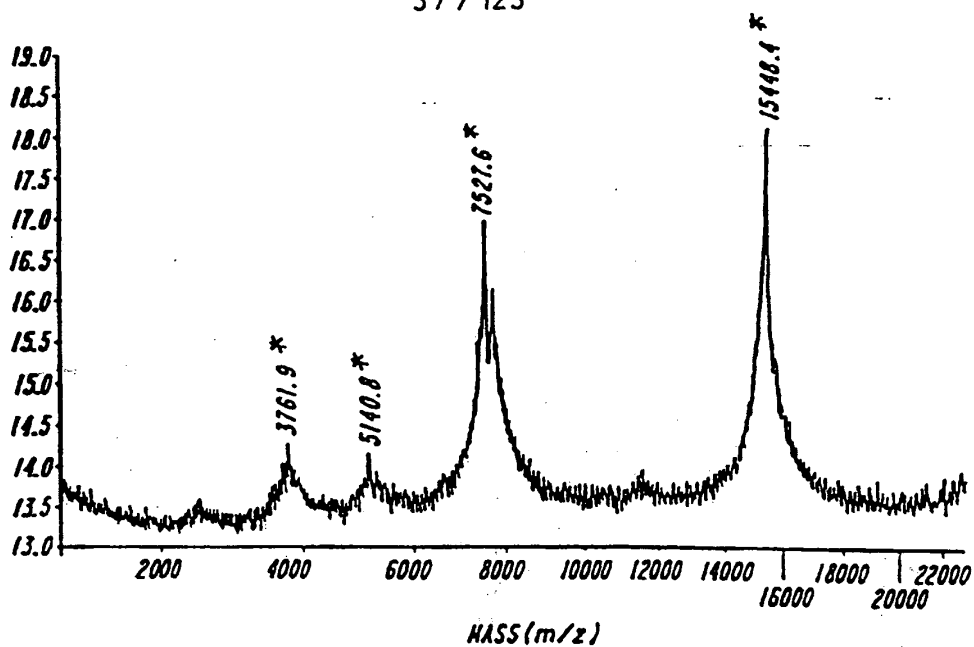
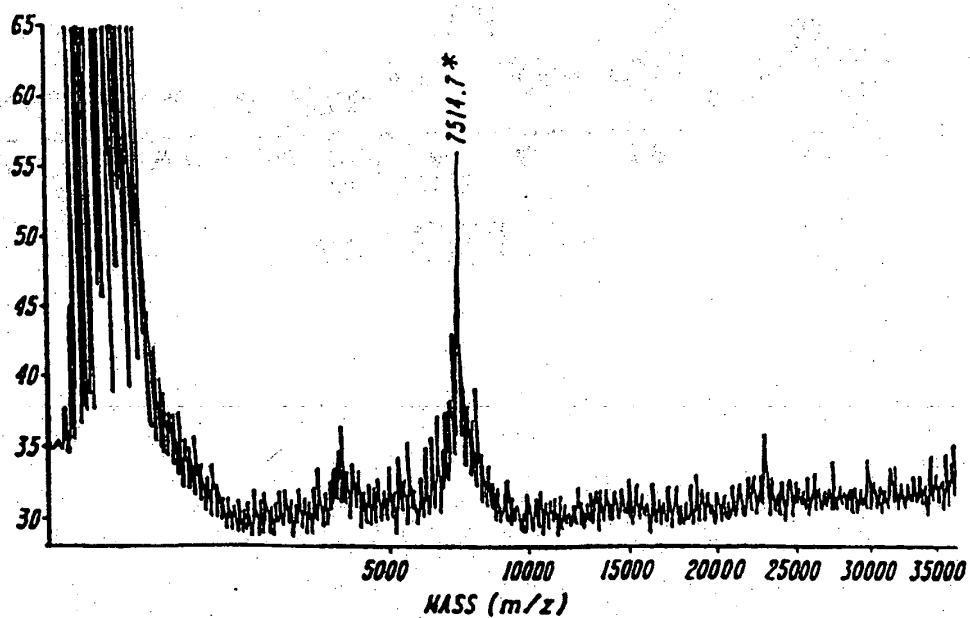


FIG. 31B

36 /123

**FIG. 32**

37 / 123

*FIG. 33A**FIG. 33B*

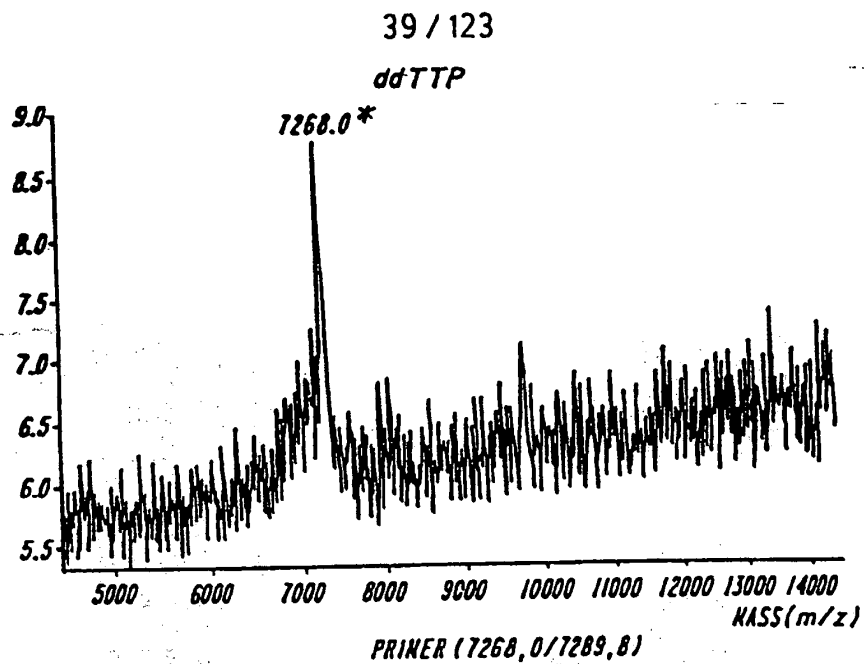
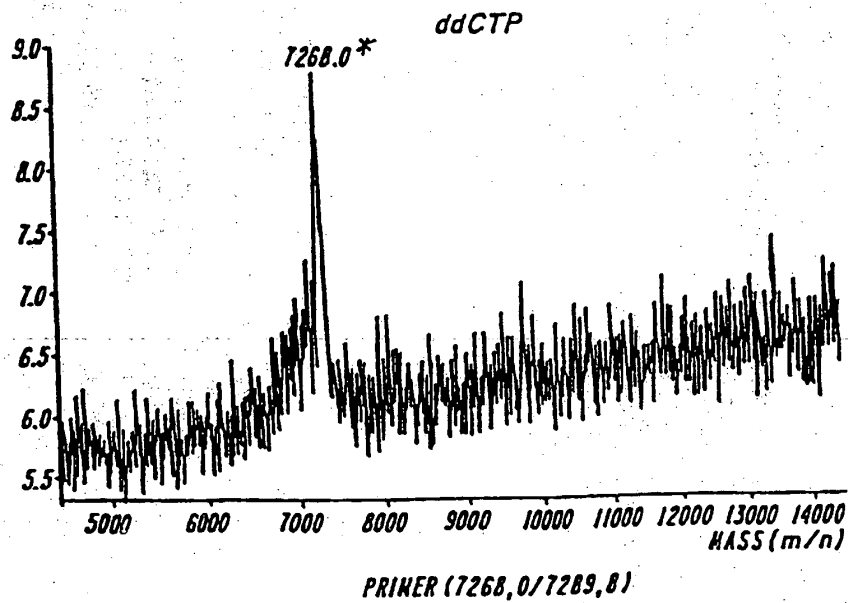
38/123

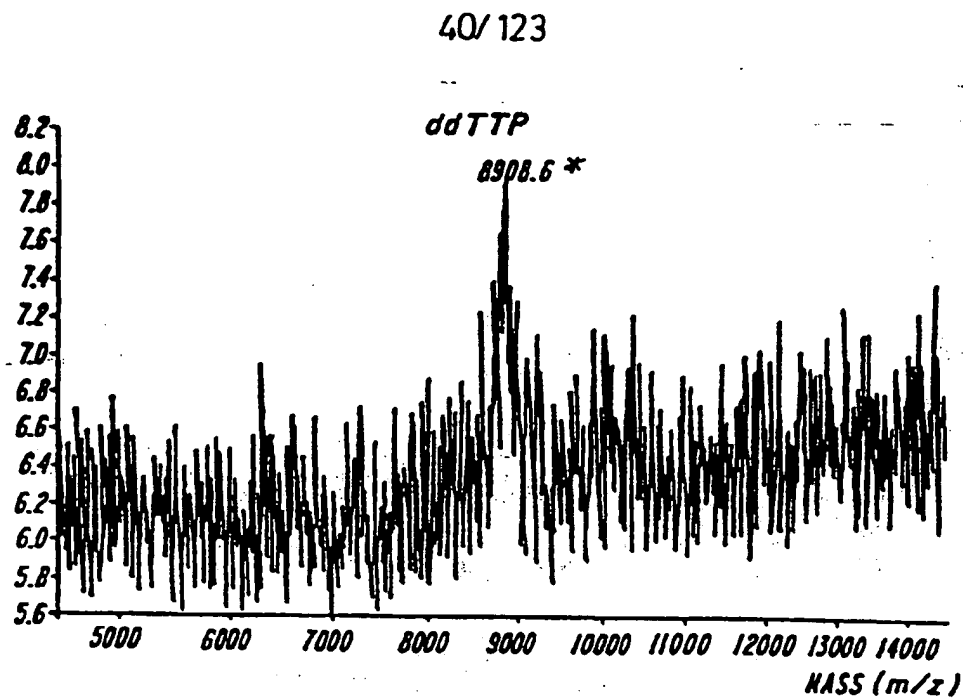
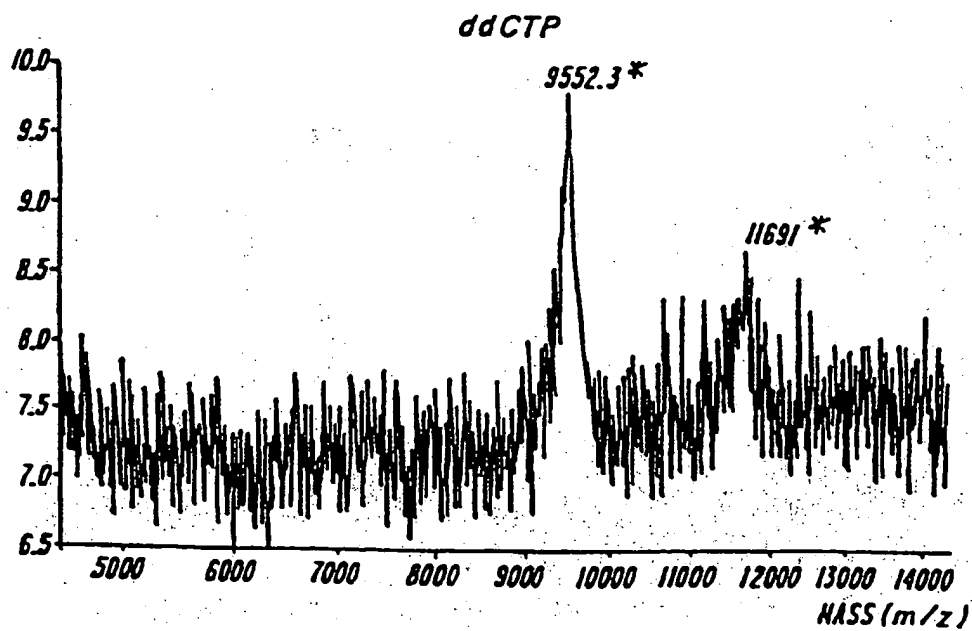
506507508  
IleIlePhe  
ACCATTAAGAAATATCATCTTTGGTGTTCCTATGATGAATATAGAACGTCATC  
primer ACCACAAAGGATACTACTTATATC (7289,8)  
wildtype TAGAAACCCACAAAGGATACTACTTATATC (8846,8)  
ΔF508 TA---ACCACAAAGGATACTACTTATATC (7891,2)  
ΔI507 TAG---AAACCACAAAGGATACTACTTATATC (8846,8)

FIG. 34A

506507508  
IleIlePhe  
ACCATTAAGAAATATCATCTTTGGTGTTCCTATGATGAATATAGAACGTCATC  
primer ACCACAAAGGATACTACTTATATC (7289,8)  
wildtype CTTTATAGTAGAAACCCACAAAGGATACTACTTATATC (11612,6)  
ΔF508 CTTTATAGTA---ACCACAAAGGATACTACTTATATC (10657,0)  
ΔI507 CTTTATAG---AAACCACAAAGGATACTACTTATATC (10666,0)  
506Ser CGTAGAAACCCACAAAGGATACTACTTATATC (9465,2)

FIG. 34B

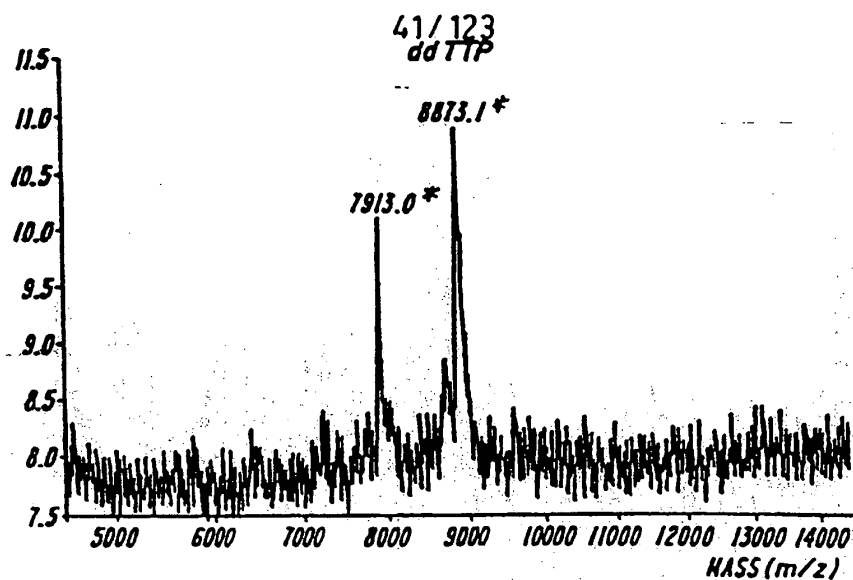
**FIG. 35A****FIG. 35B**

**FIG. 35C**

5065 (9552, 3 / 9465, 2)

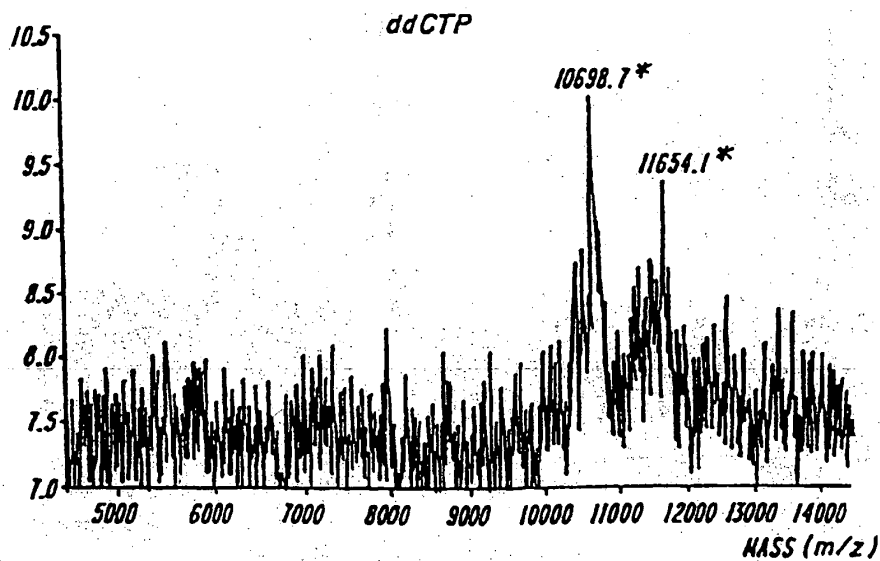
WILDTYPE (11691, 9 / 11612, 6)

**FIG. 35D**



ΔF508 (7913,0/7891,2)  
WILDTYPE (8873,1/8846,8)

**FIG. 35E**



Δ508 (10698,7/10657,0)  
WILDTYPE (11654,1/11612,6)

**FIG. 35F**



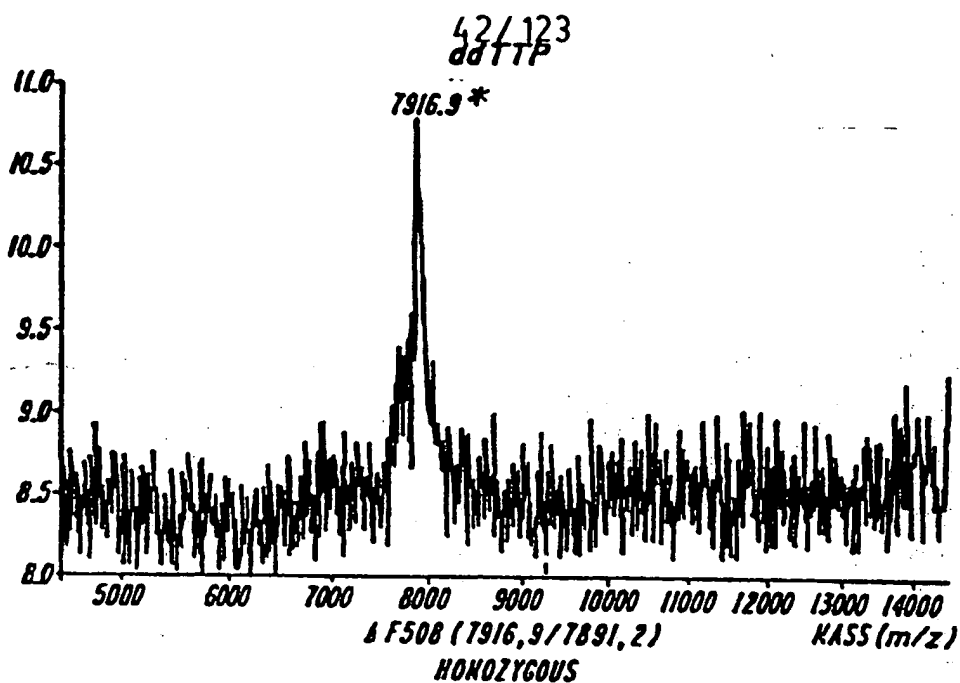


FIG. 35G

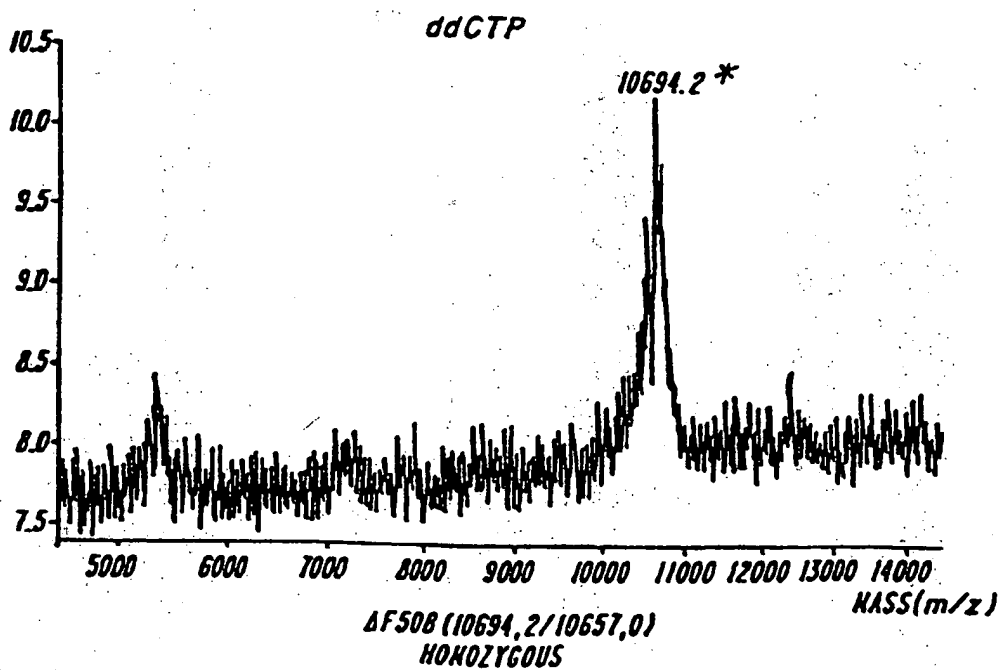
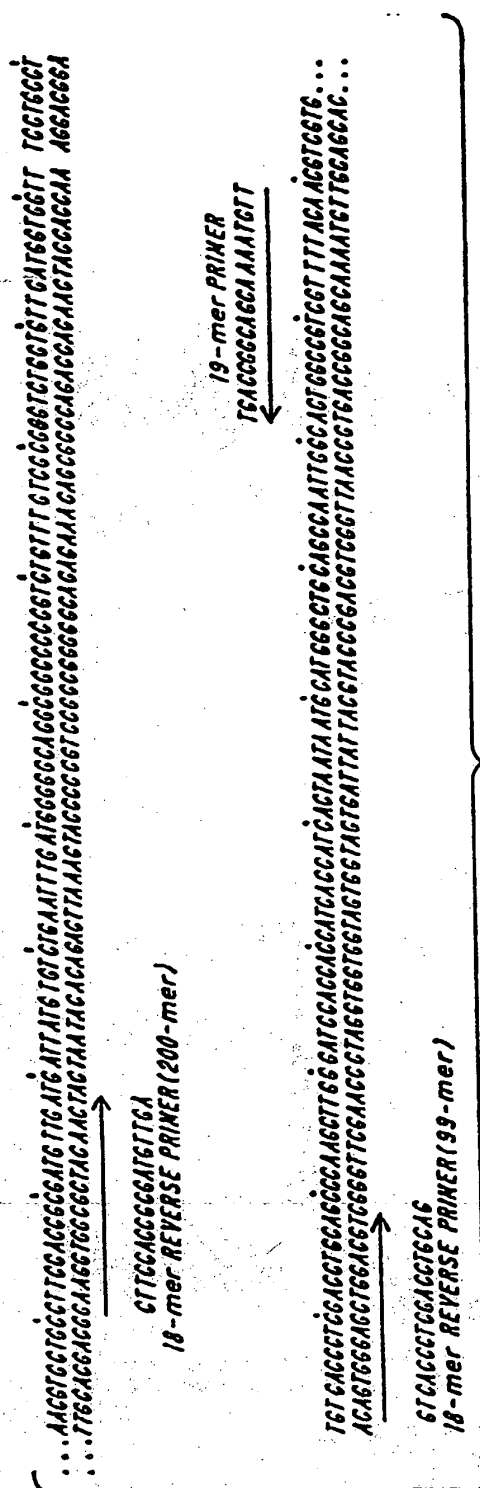
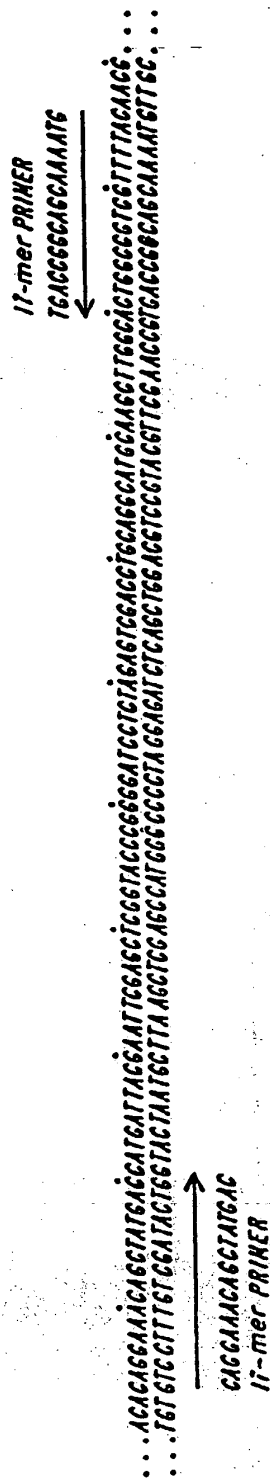


FIG. 35H

43 / 123

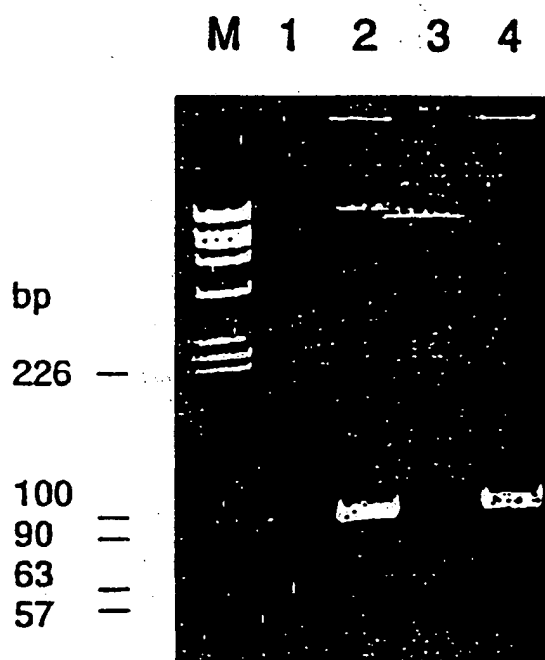


**FIG. 36**

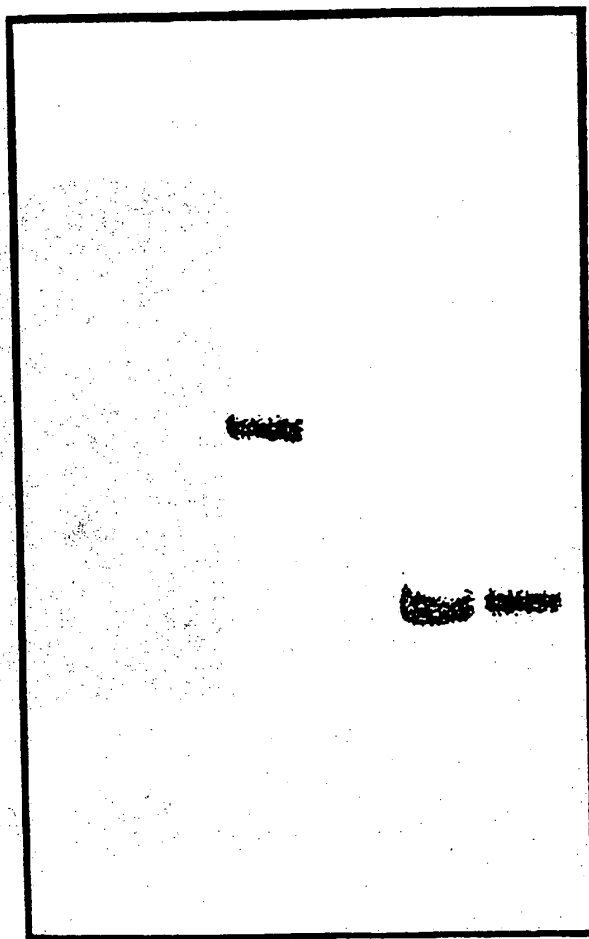


**FIG. 37**

44 / 123

*FIG. 38*

1 2 3 4 5 6



**FIG. 39**

46 / 123

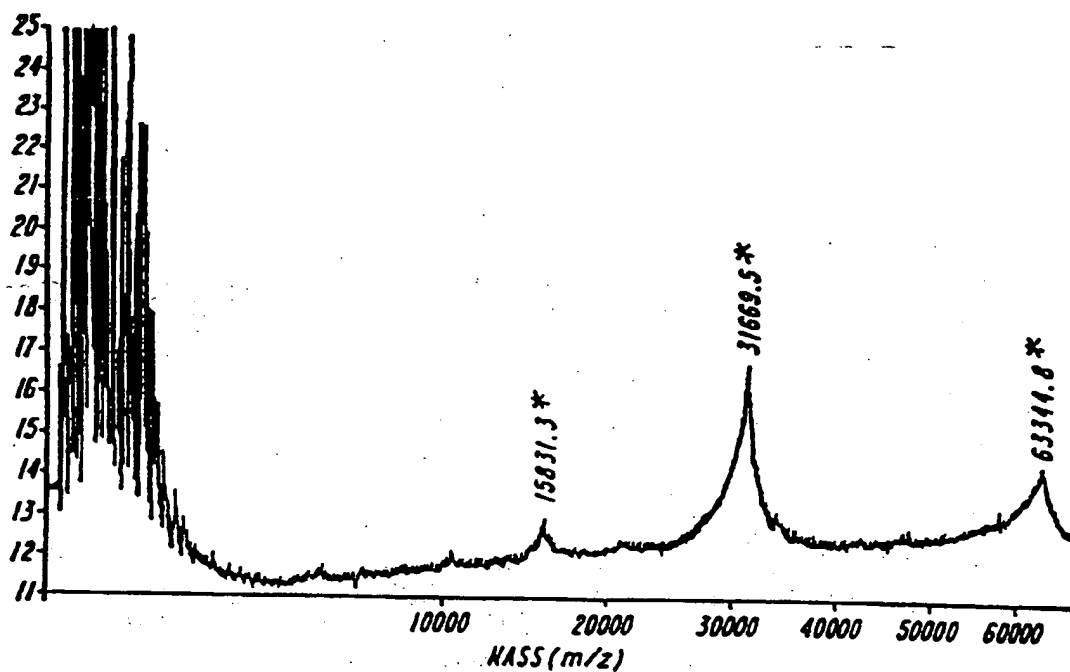


FIG. 40A

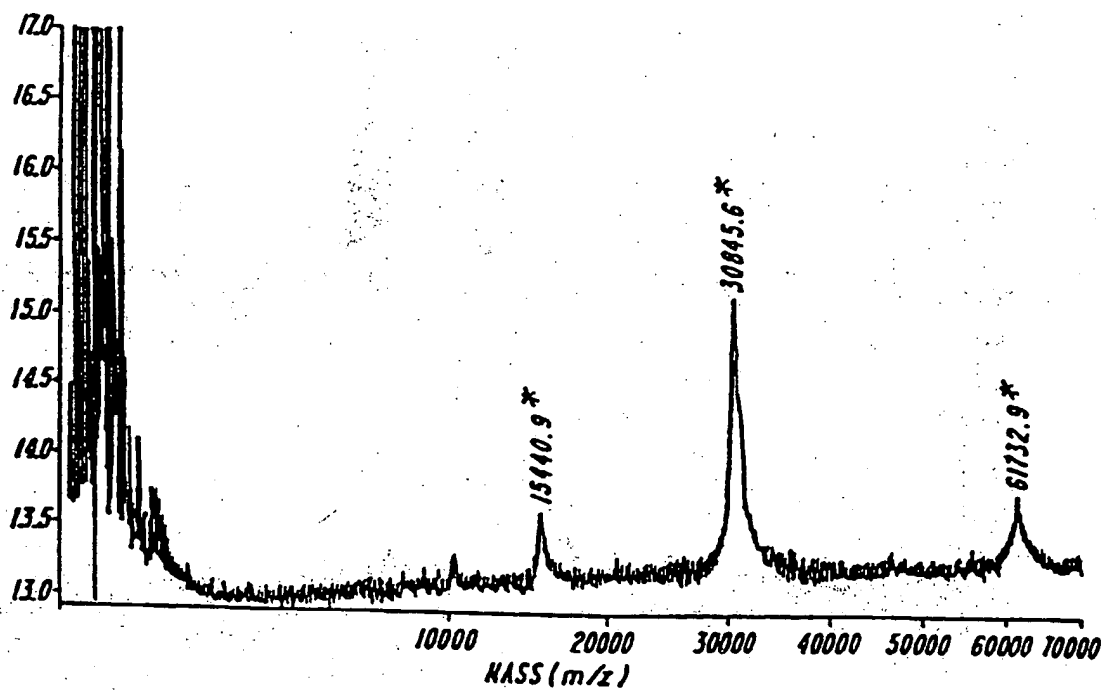


FIG. 40B

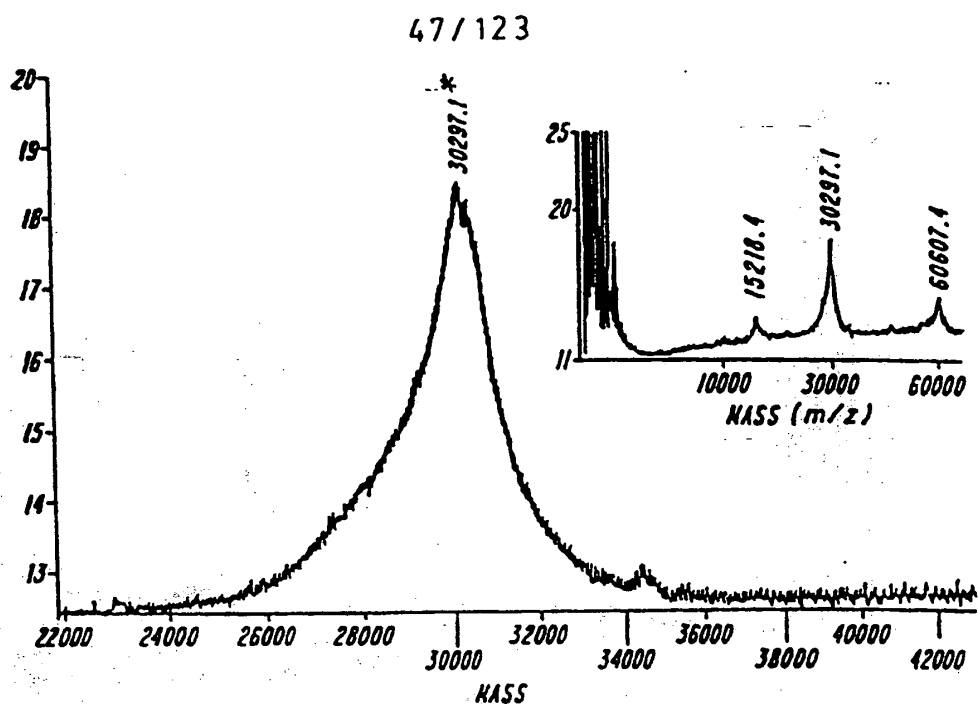


FIG. 41A

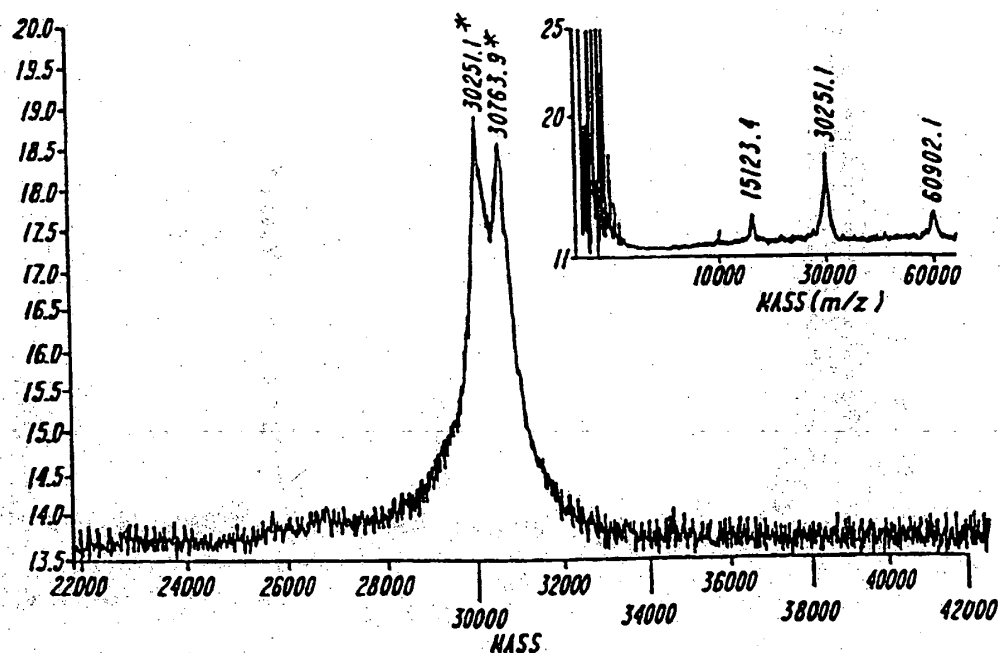


FIG. 41B

48/123

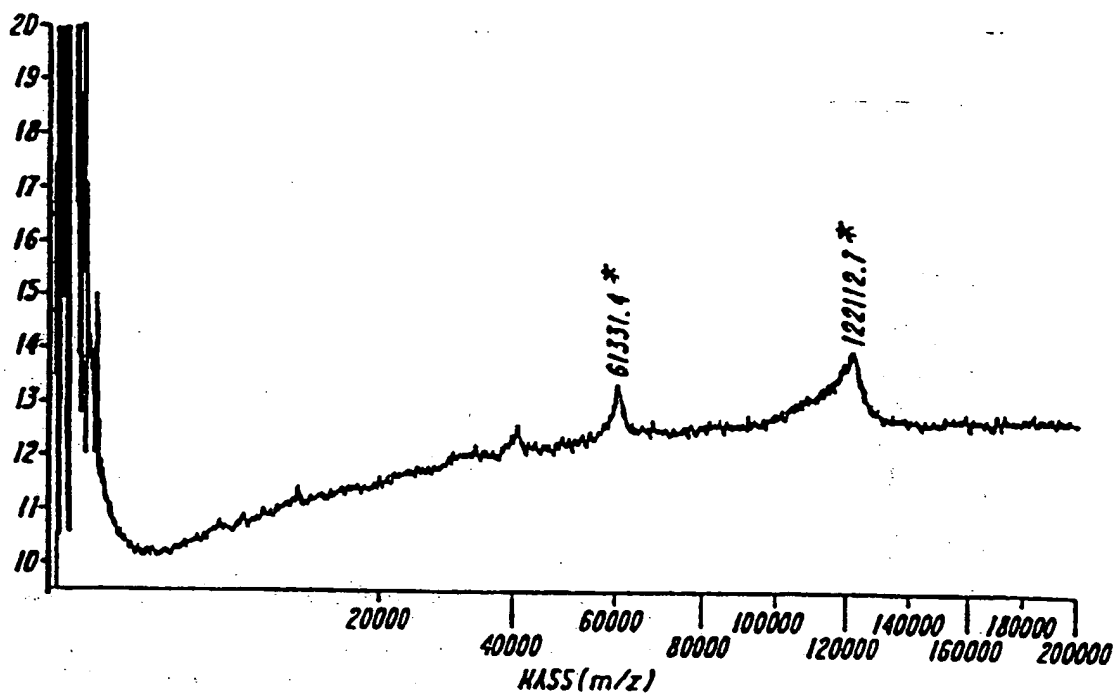


FIG. 42A

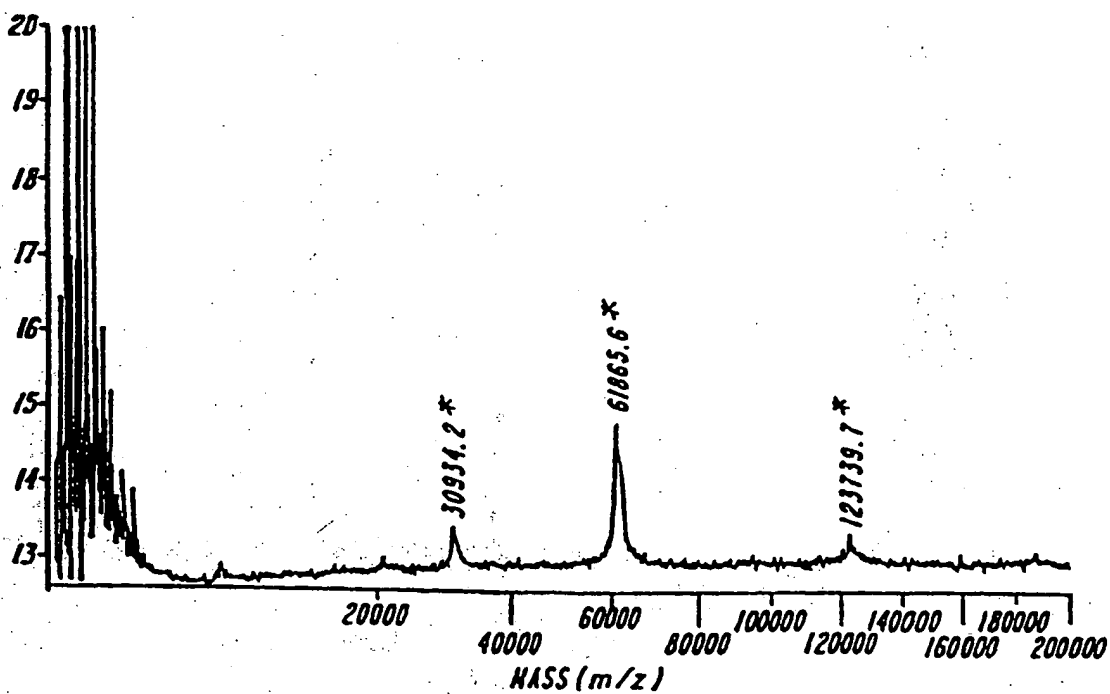


FIG. 42B

49 / 123

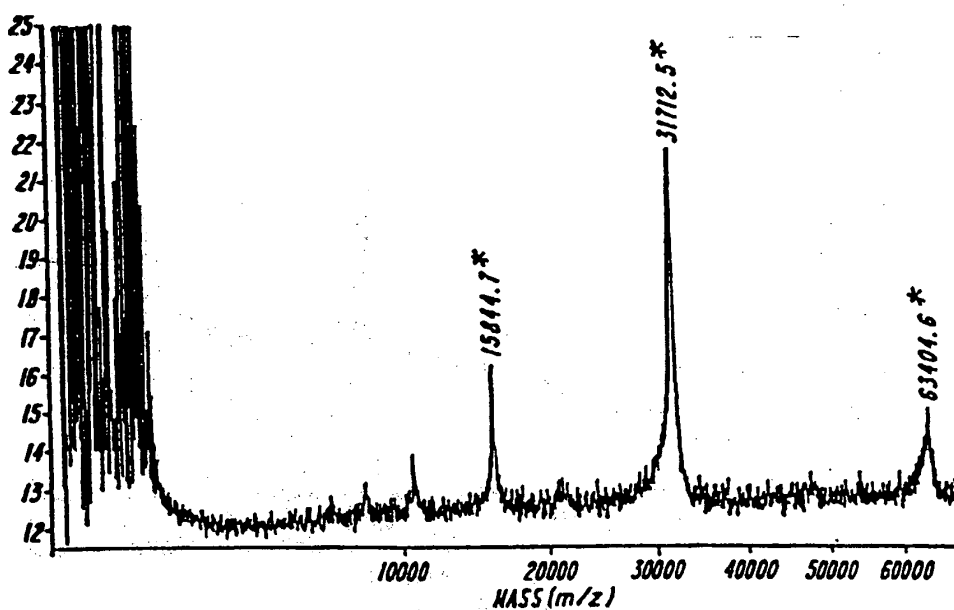


FIG. 43A

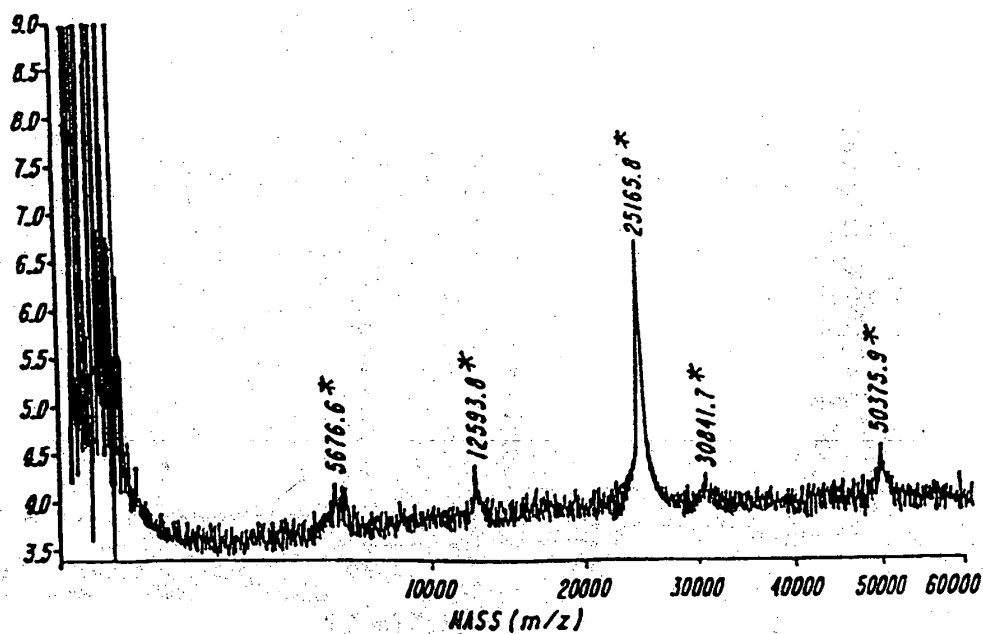


FIG. 43B



50 / 123

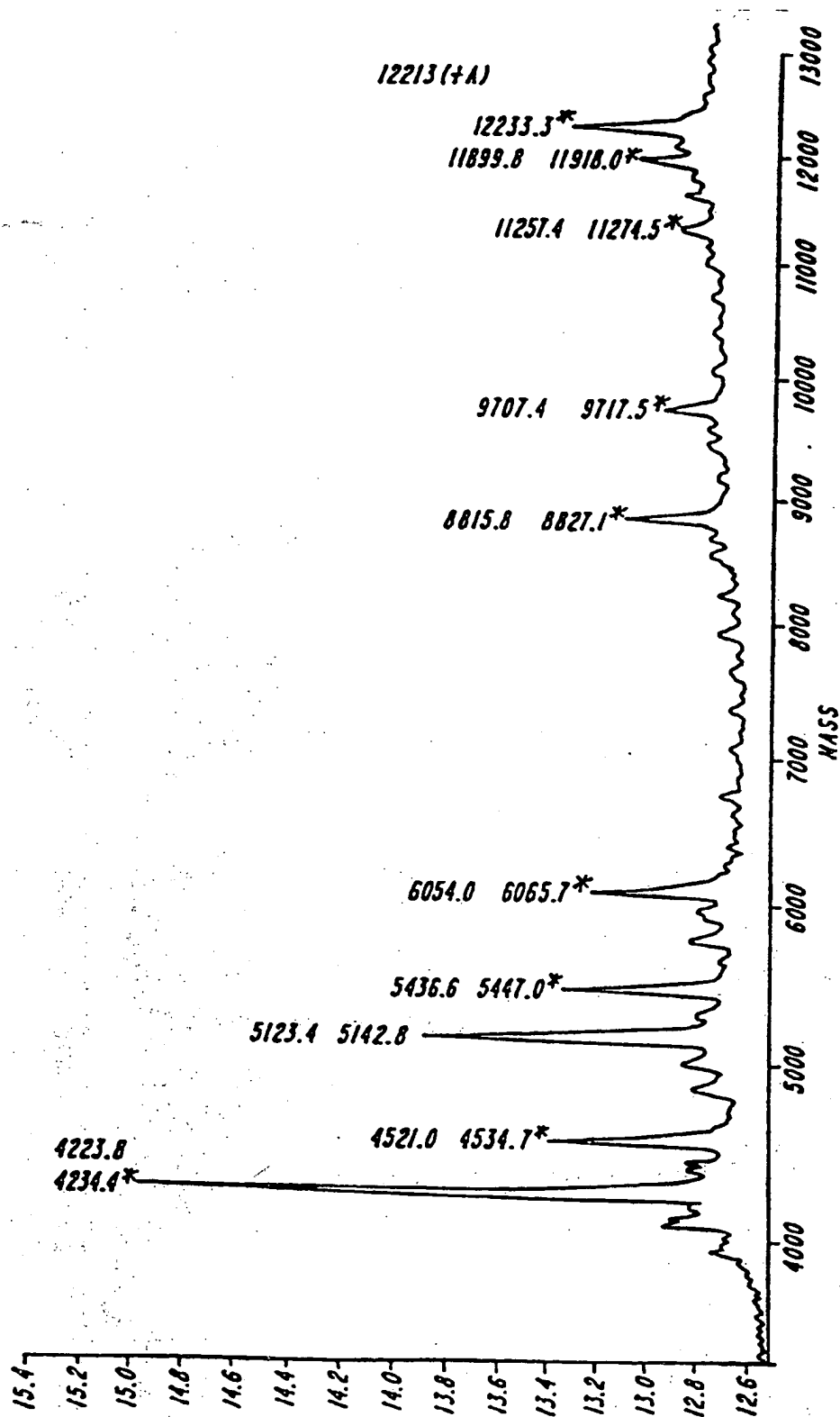


FIG. 44A

51 / 123

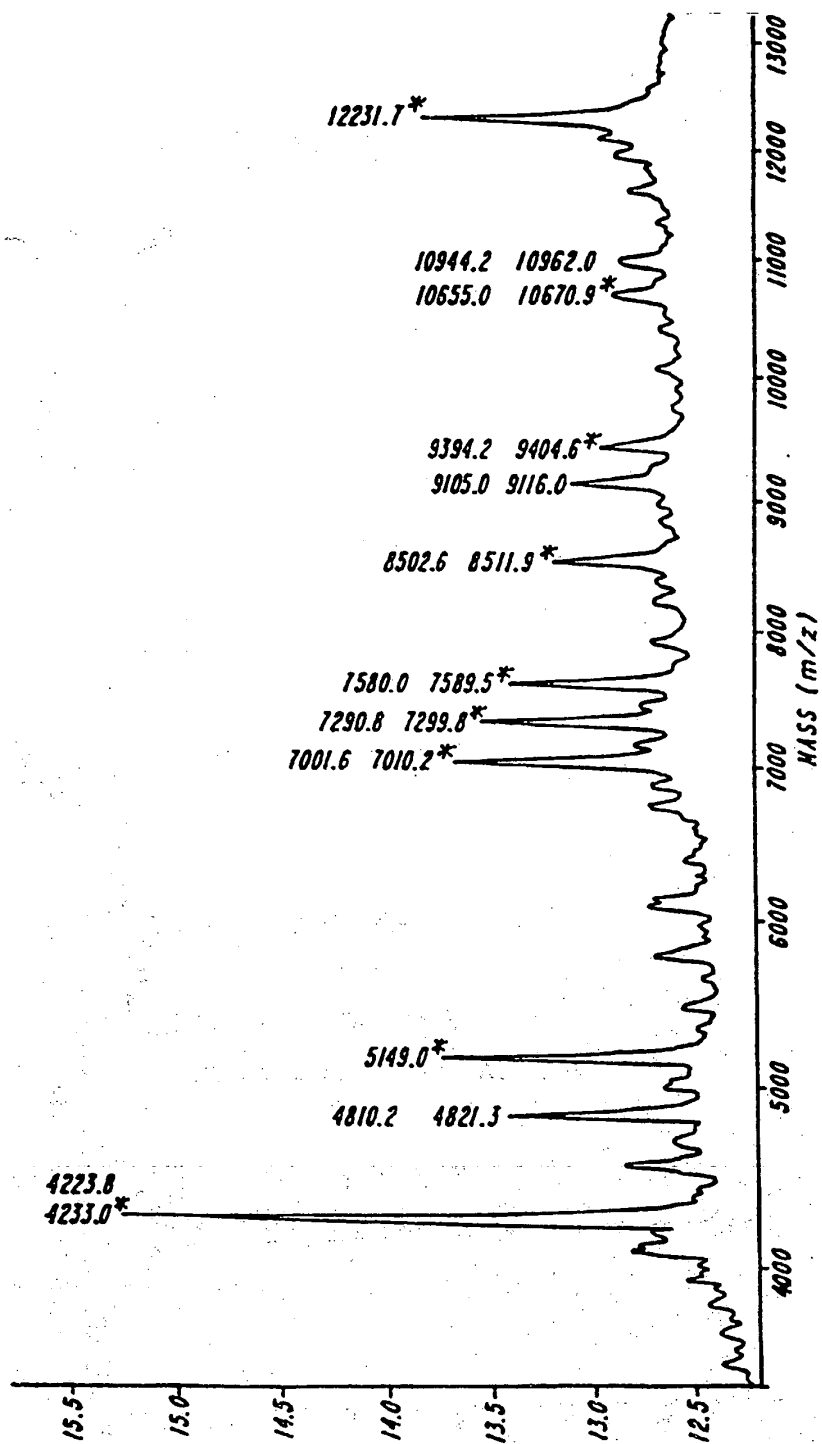


FIG. 44B

52/123

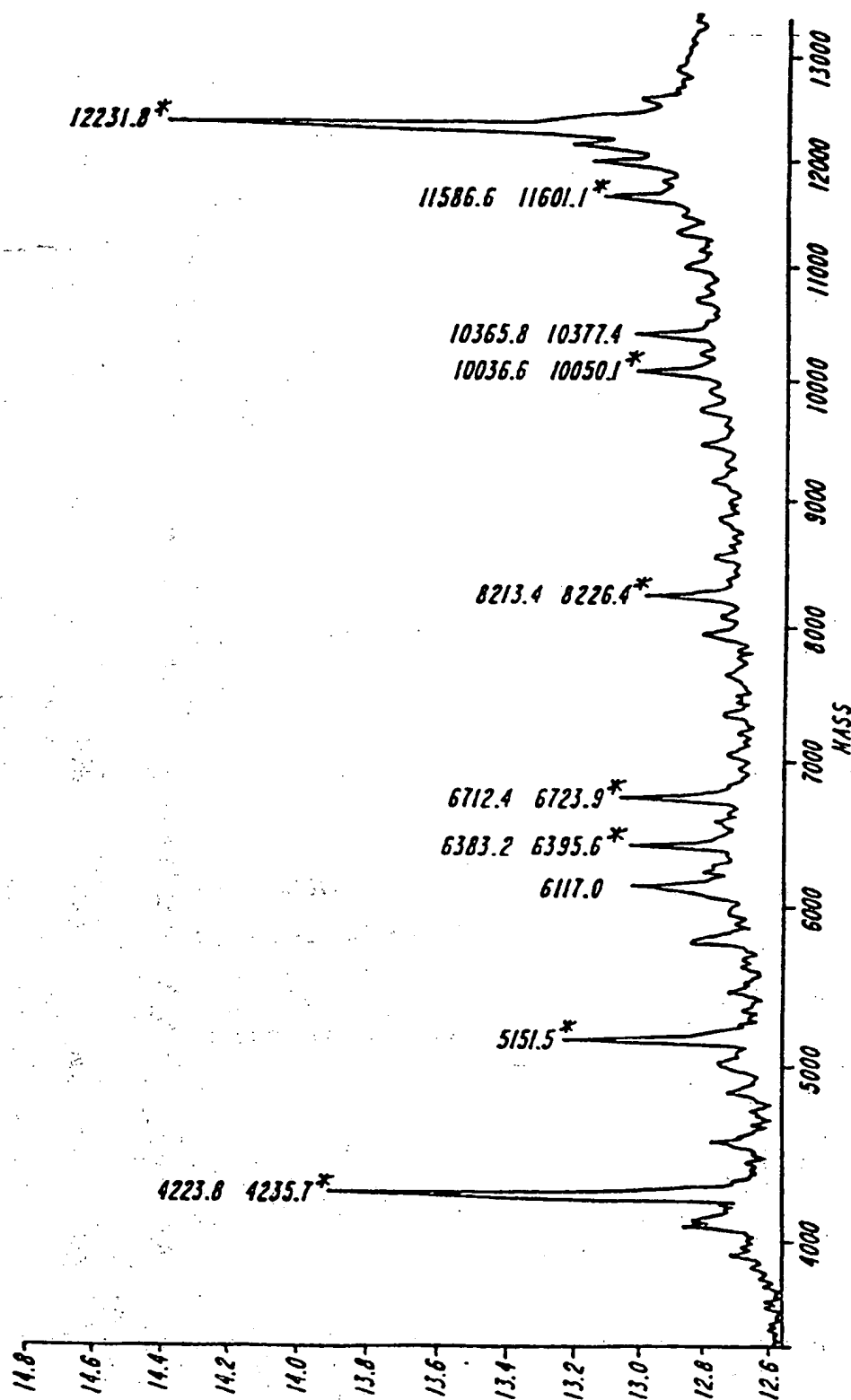


FIG. 44C

53/123

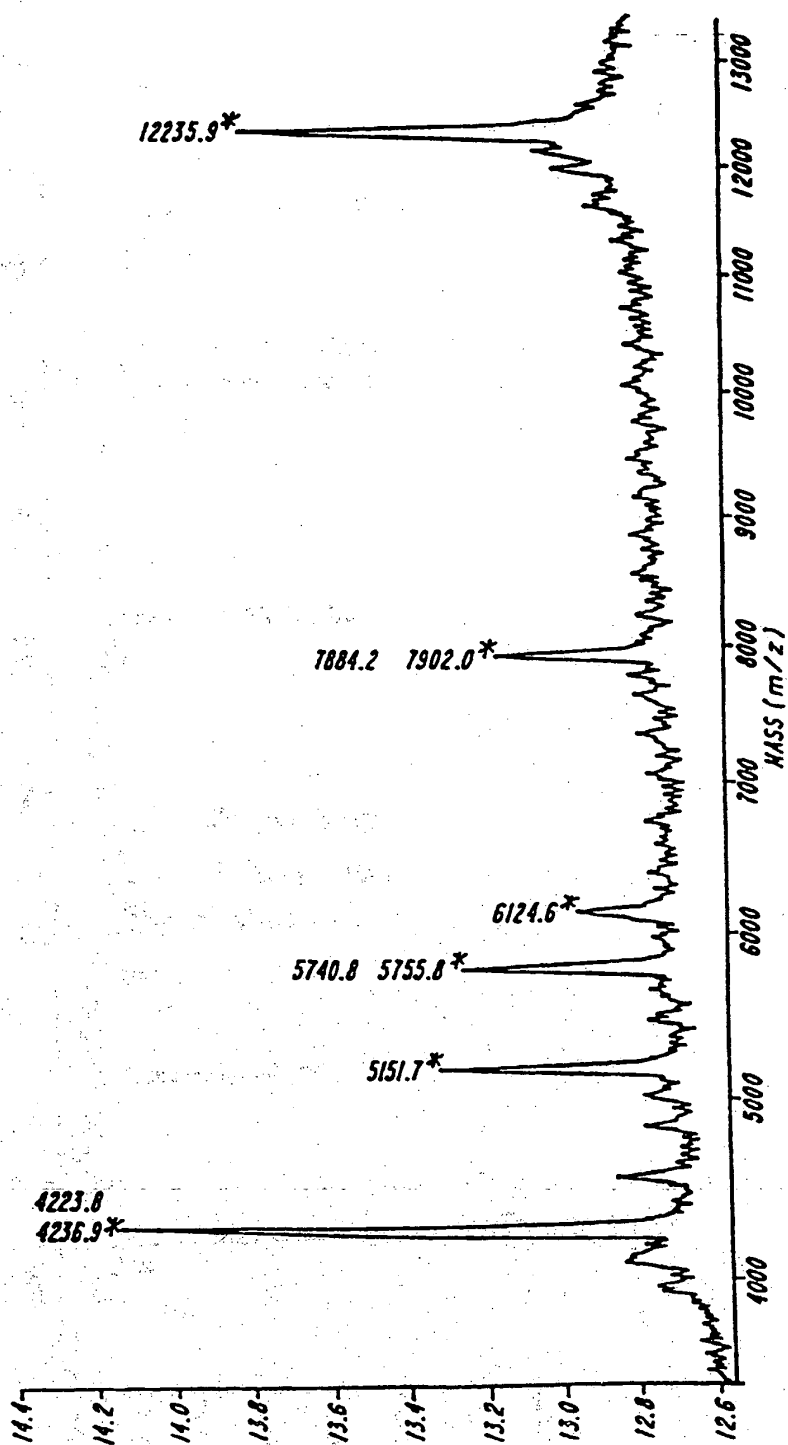


FIG. 44D

54 / 123

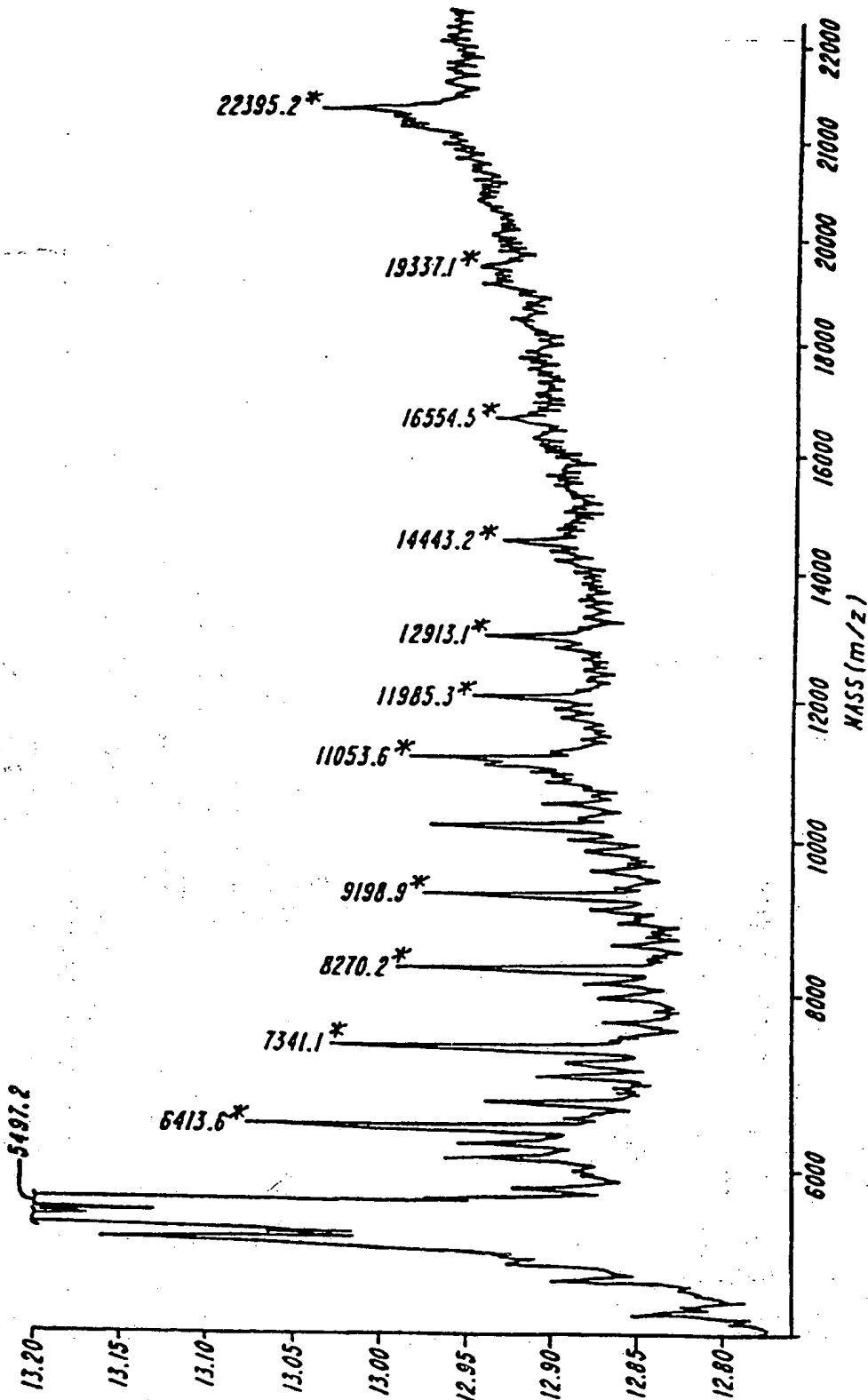


FIG. 45

55 / 123

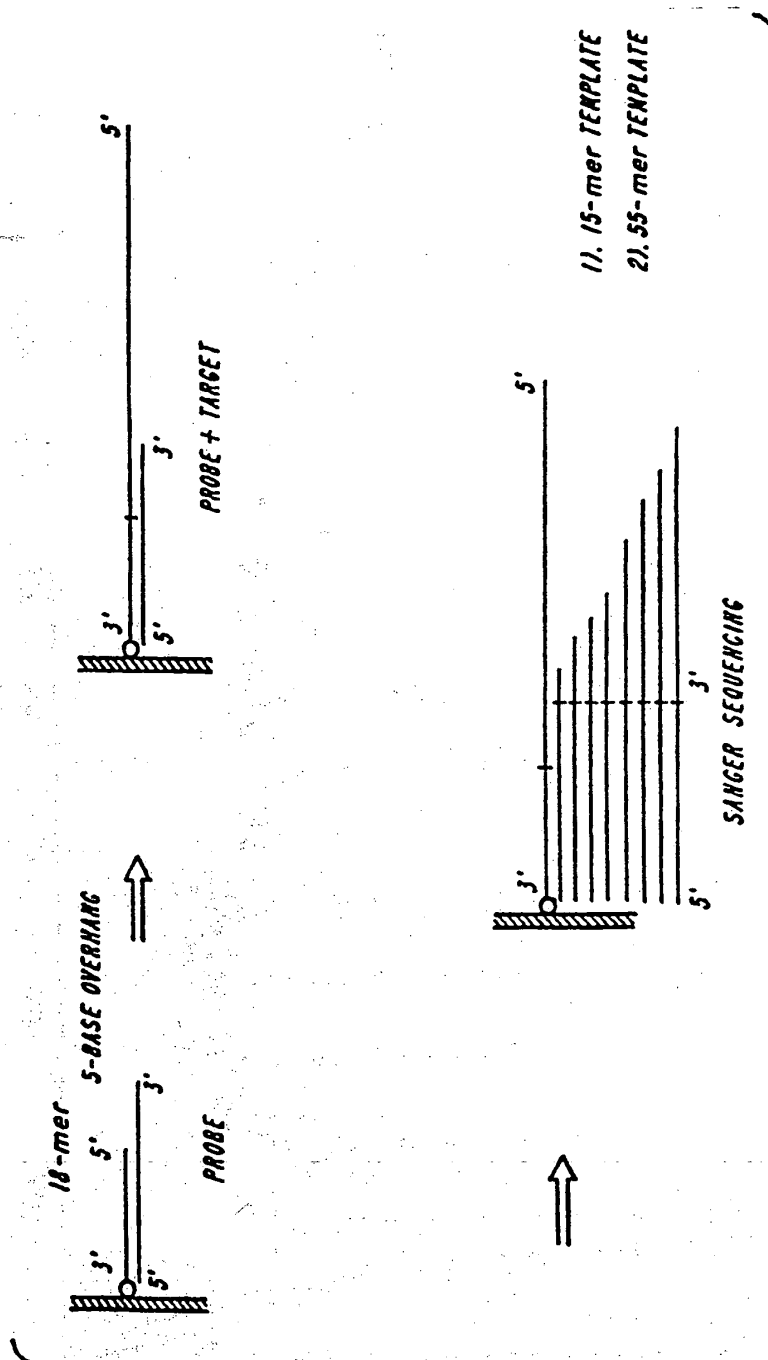


FIG. 46

56 / 123

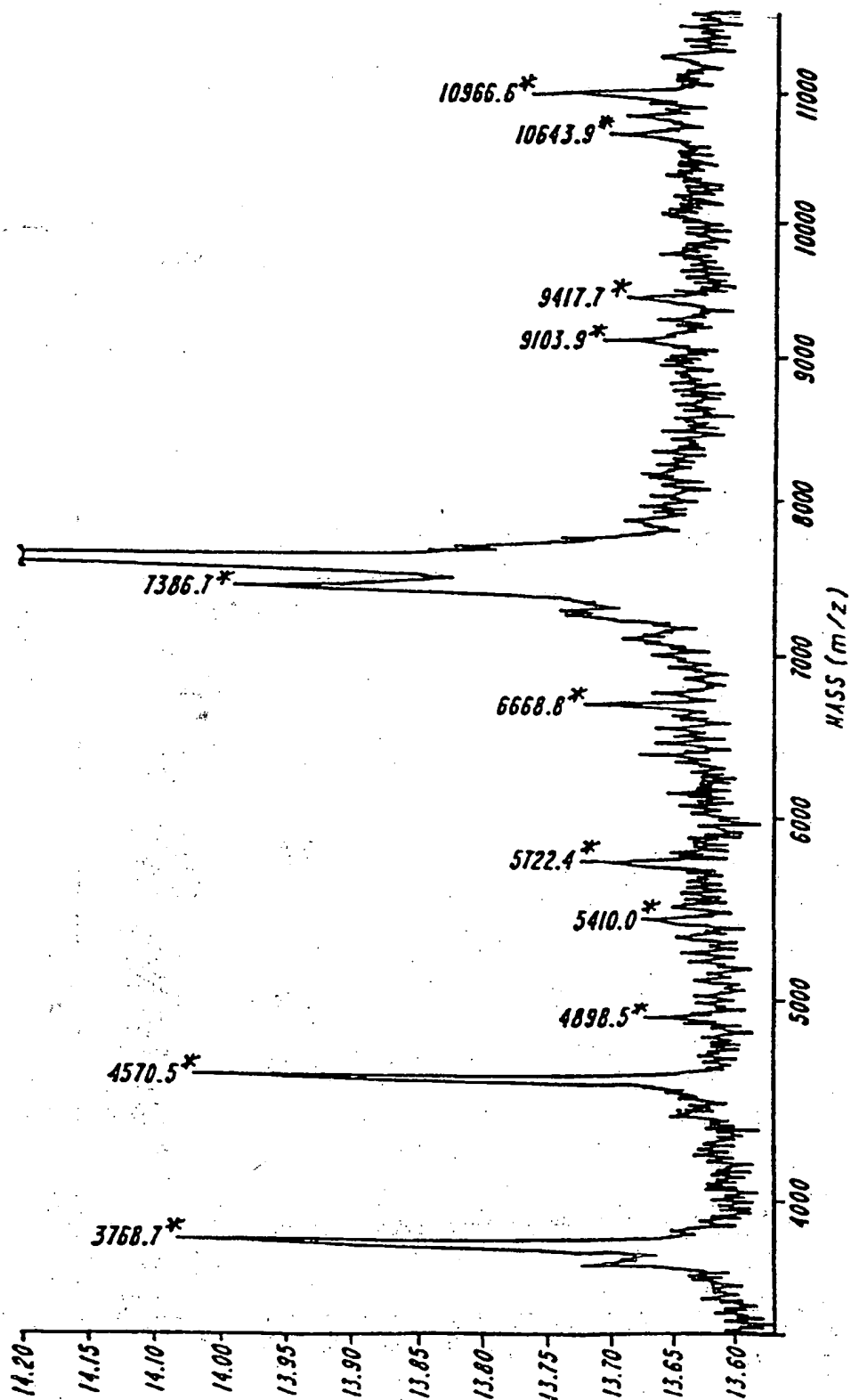


FIG. 47A

57/123

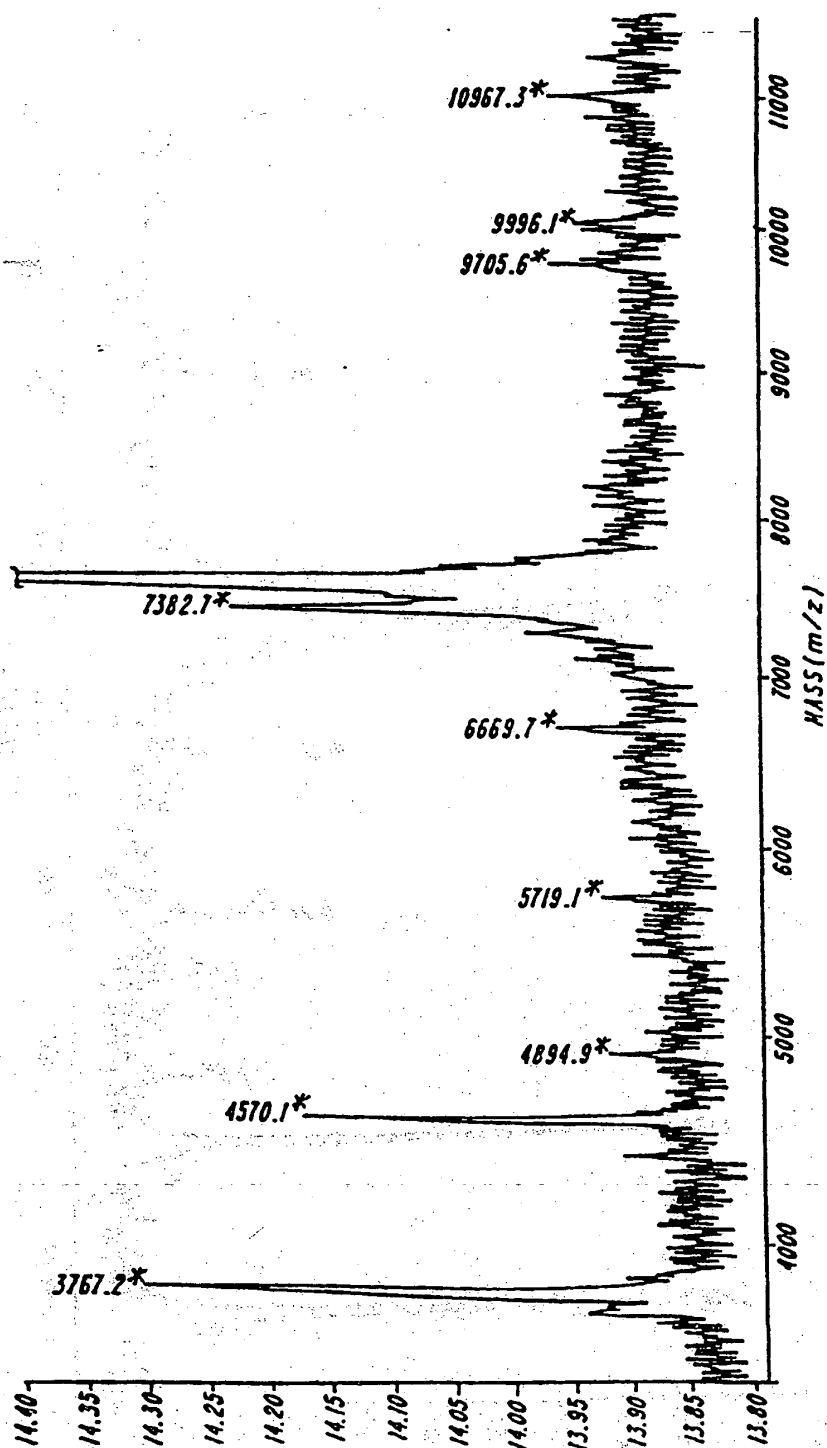


FIG. 47B



58/123

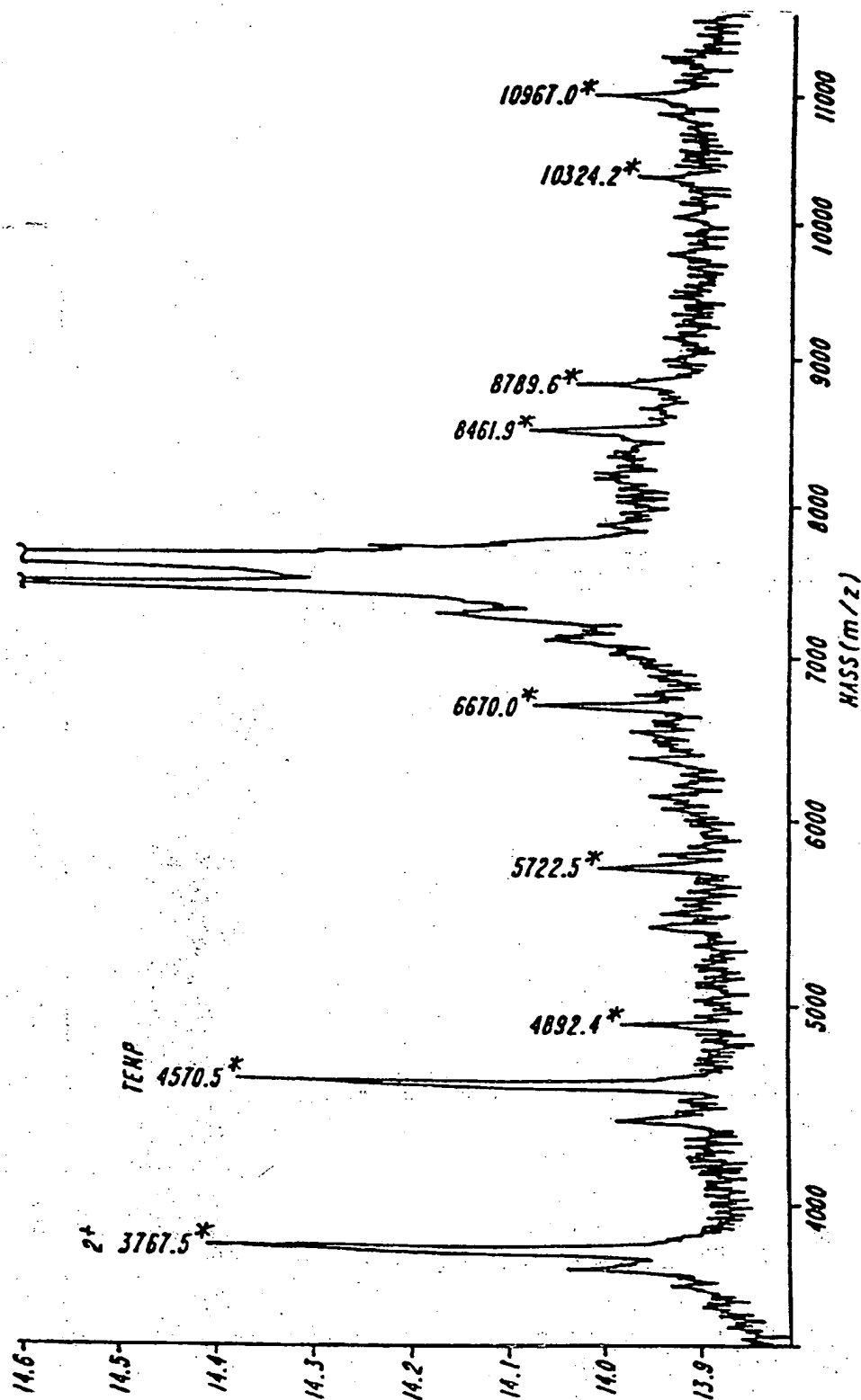


FIG. 47C

59/123

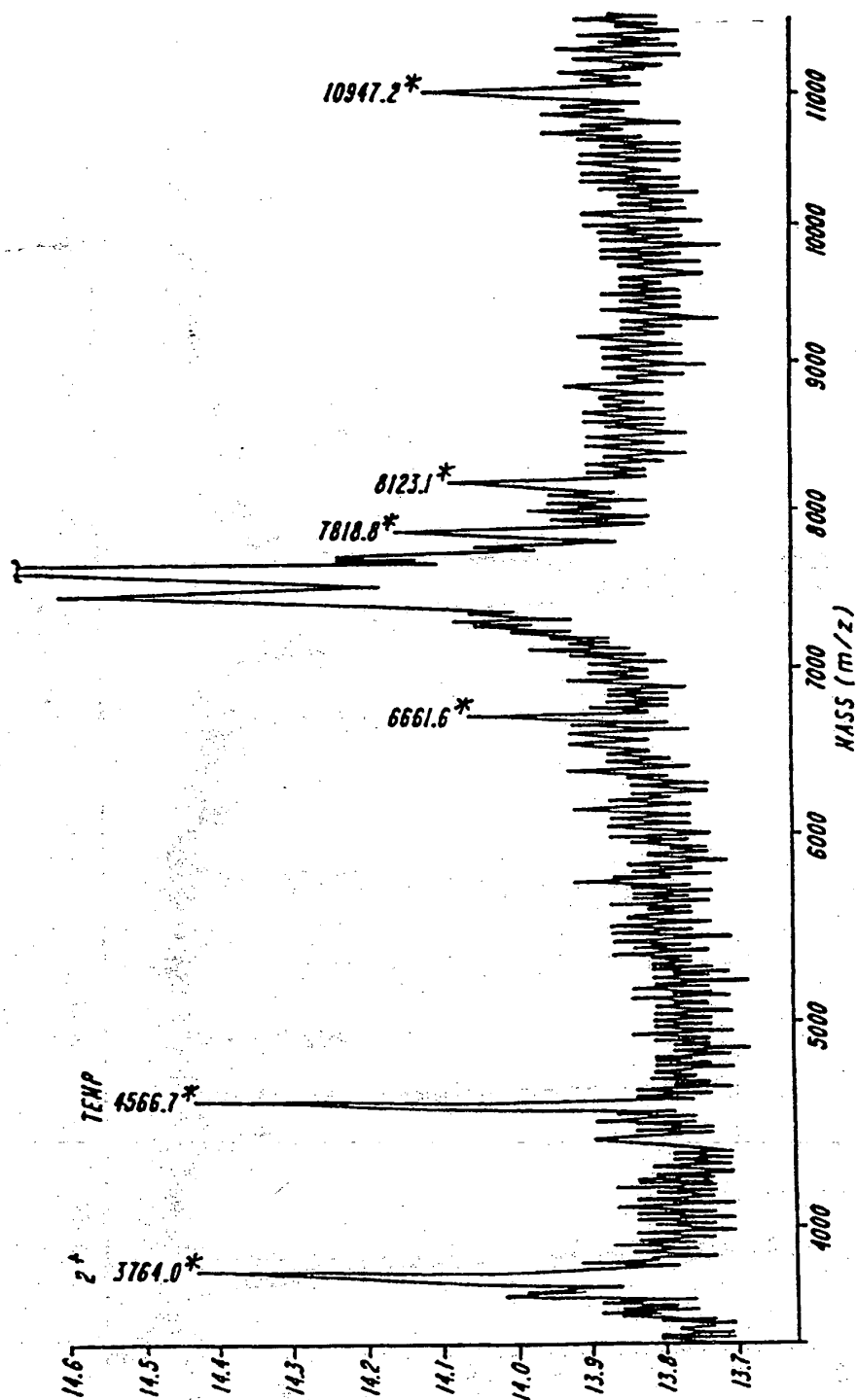


FIG. 47D

60/123

FIG. 48A

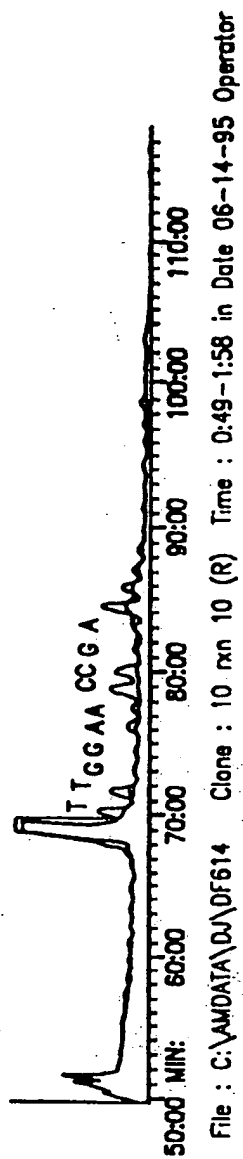


FIG. 48B

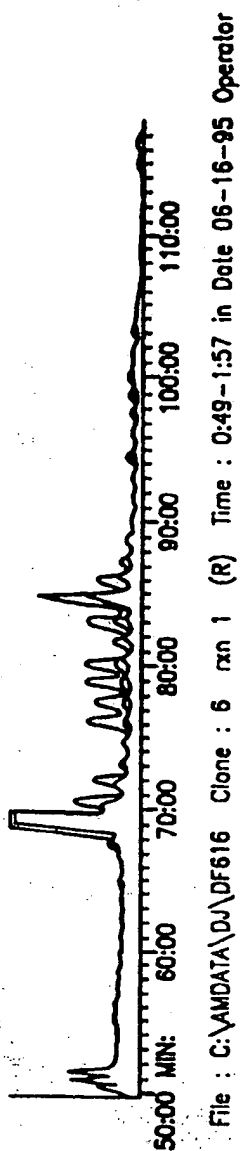


FIG. 48C

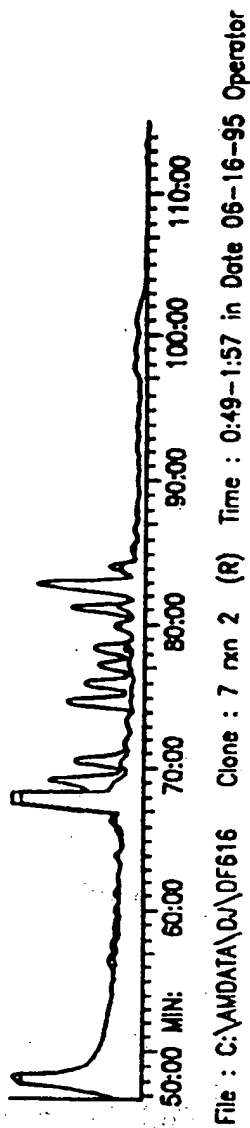
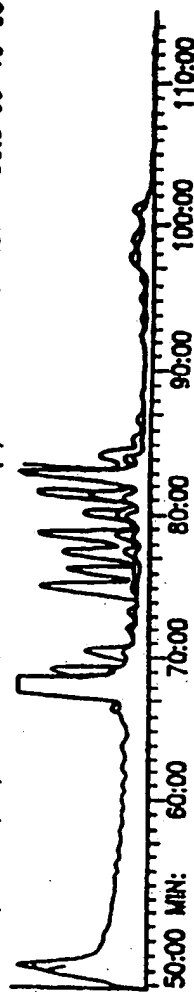


FIG. 48D



61/123

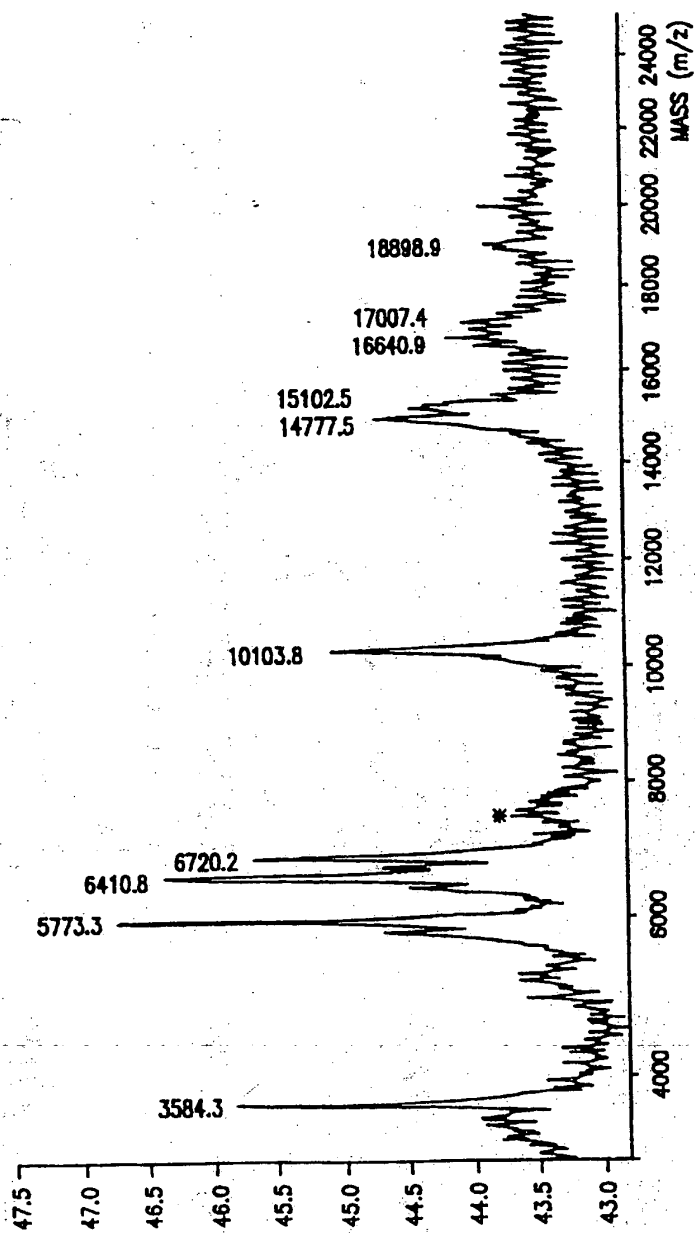


FIG. 49A

62 / 123

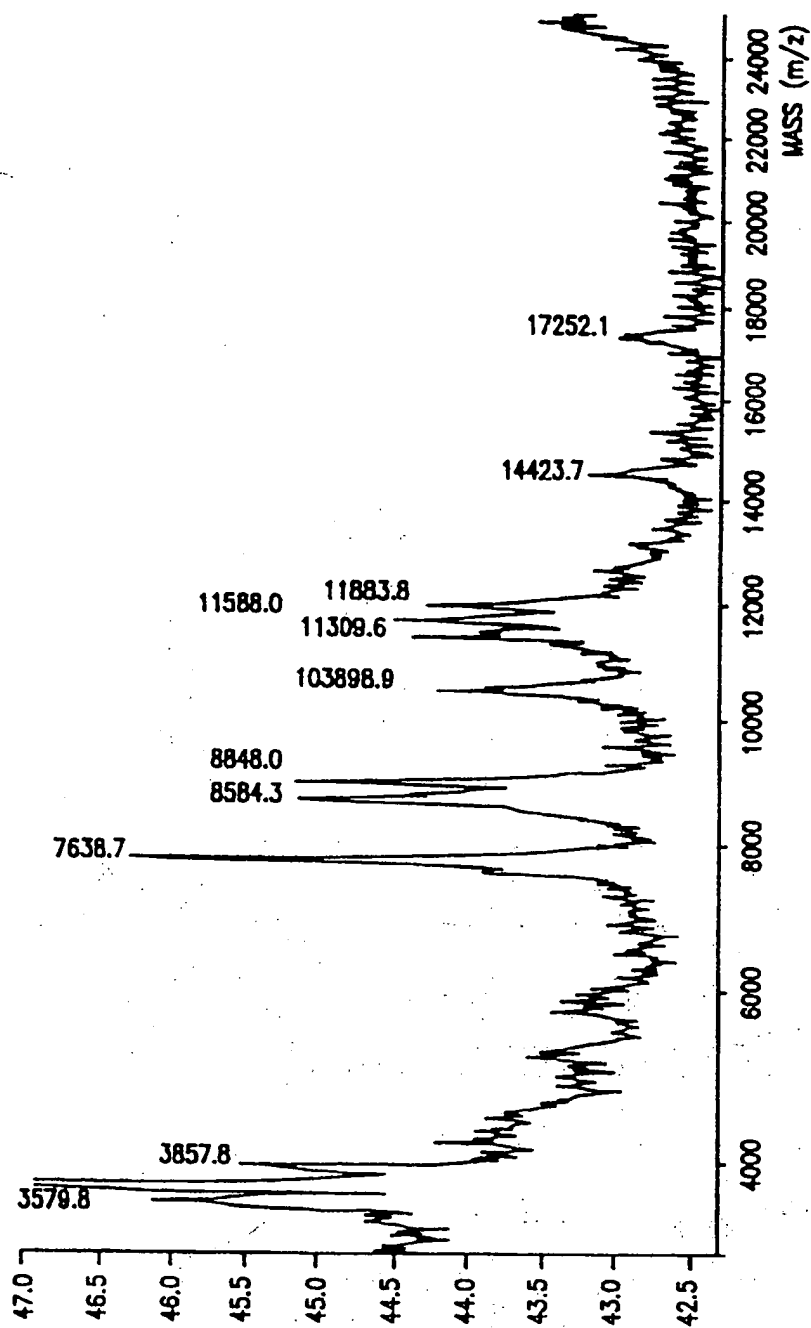


FIG. 49B

63 / 123

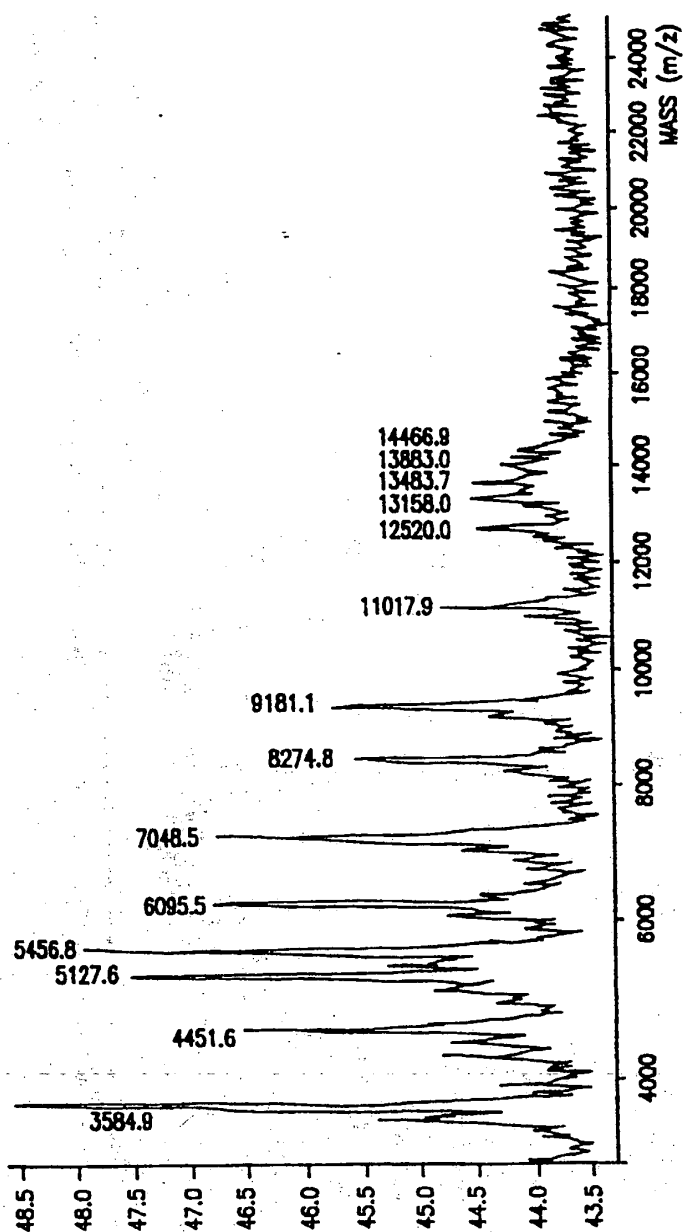


FIG. 49C

64/123

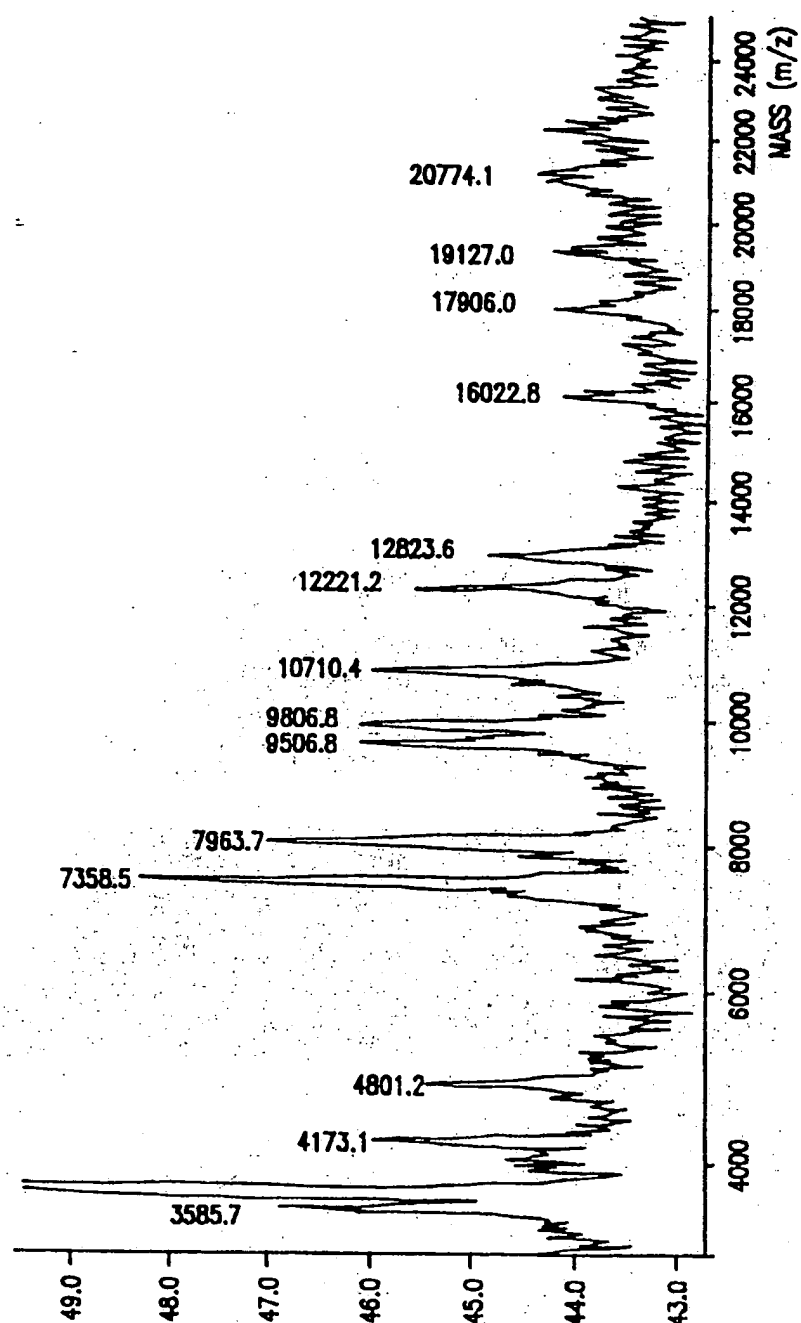


FIG. 49D

65/123

3'-(H)<sub>n</sub>-ACCACGTGGACTGAG GACACCTCTT CAGACGGCAA TGACGOGACA CCCCCTTCCA CTTCACCTA-(N)<sub>n</sub>-5'

PARTIAL SEQUENCE OF THE  $\beta$ -GLOBIN TEMPLATE

5-TGCACCTGACTC	3' (PRIME)
5-TGCACCTGACTC	C-3'
5-TGCACCTGACTC	CT-3'
5-TGCACCTGACTC	CTC-3'
5-TGCACCTGACTC	CTGT-3'
5-TGCACCTGACTC	CTGTG-3'
5-TGCACCTGACTC	CTGTGG-3'
5-TGCACCTGACTC	CTGTGGA-3'
5-TGCACCTGACTC	CTGTGGAG-3'
5-TGCACCTGACTC	CTGTGGAGA-3'
5-TGCACCTGACTC	CTGTGGAGAA-3'
5-TGCACCTGACTC	CTGTGGAGAA G-3'
5-TGCACCTGACTC	CTGTGGAGAA GT-3'
5-TGCACCTGACTC	CTGTGGAGAA GTC-3'
5-TGCACCTGACTC	CTGTGGAGAA GTCT-3'
5-TGCACCTGACTC	CTGTGGAGAA GTCTG-3'
5-TGCACCTGACTC	CTGTGGAGAA GTCTGC-3'
5-TGCACCTGACTC	CTGTGGAGAA GTCTGCC-3'
5-TGCACCTGACTC	CTGTGGAGAA GTCTGCCG-3'
5-TGCACCTGACTC	CTGTGGAGAA GTCTGCCGT-3'
5-TGCACCTGACTC	CTGTGGAGAA GTCTGCCGTT-3'
5-TGCACCTGACTC	CTGTGGAGAA GTCTGCCGTT A-3'
5-TGCACCTGACTC	CTGTGGAGAA GTCTGCCGTT AC-3'
5-TGCACCTGACTC	CTGTGGAGAA GTCTGCCGTT ACT-3'
5-TGCACCTGACTC	CTGTGGAGAA GTCTGCCGTT ACTG-3'
5-TGCACCTGACTC	CTGTGGAGAA GTCTGCCGTT ACTGC-3'
5-TGCACCTGACTC	CTGTGGAGAA GTCTGCCGTT ACTGCC-3'
5-TGCACCTGACTC	CTGTGGAGAA GTCTGCCGTT ACTGCCC-3'
5-TGCACCTGACTC	CTGTGGAGAA GTCTGCCGTT ACTGCCCT-3'
5-TGCACCTGACTC	CTGTGGAGAA GTCTGCCGTT ACTGCCCTG-3'
5-TGCACCTGACTC	CTGTGGAGAA GTCTGCCGTT ACTGCCCTGT-3'
5-TGCACCTGACTC	CTGTGGAGAA GTCTGCCGTT ACTGCCCTGT G-3'
5-TGCACCTGACTC	CTGTGGAGAA GTCTGCCGTT ACTGCCCTGT GG-3'
5-TGCACCTGACTC	CTGTGGAGAA GTCTGCCGTT ACTGCCCTGT GGG-3'
5-TGCACCTGACTC	CTGTGGAGAA GTCTGCCGTT ACTGCCCTGT GGGG-3'
5-TGCACCTGACTC	CTGTGGAGAA GTCTGCCGTT ACTGCCCTGT GGGGC-3'
5-TGCACCTGACTC	CTGTGGAGAA GTCTGCCGTT ACTGCCCTGT GGGGCA-3'
5-TGCACCTGACTC	CTGTGGAGAA GTCTGCCGTT ACTGCCCTGT GGGGCAA-3'
5-TGCACCTGACTC	CTGTGGAGAA GTCTGCCGTT ACTGCCCTGT GGGGCAAG-3'
5-TGCACCTGACTC	CTGTGGAGAA GTCTGCCGTT ACTGCCCTGT GGGGCAAGG-3'
5-TGCACCTGACTC	CTGTGGAGAA GTCTGCCGTT ACTGCCCTGT GGGGCAAGGT-3'

12 met PRIMER

FIG. 50A



66/123

## REACTION STOPPED WITH

ddATP	ddCTP	ddGTP	ddTTP
3581.4 da	3581.4 da 3854.6 da	3581.4 da	3581.4 da
		4488.0 da	4158.9 da
		5120.4 da	4781.2 da
		5448.6 da	
5760.8 da		6089.0 da	
6401.2 da			
6713.4 da		7041.6 da	7344.8 da
	7634.0 da		7938.2 da
	8555.6 da	8267.4 da	
	8844.8 da	9174.0 da	9477.2 da 9781.4 da
10094.6 da			
	10382.8 da		10687.0 da
	11304.4 da	11016.2 da	
	11593.6 da		
	11652.8 da	12516.2 da	12187.0 da
			12819.4 da
		13148.6 da	
		13476.8 da	
		13805.0 da	
		14133.2 da	
	14421.4 da		
14734.6 da		15375.0 da	
15146.8 da		15703.2 da	
			16006.4 da

FIG. 50B

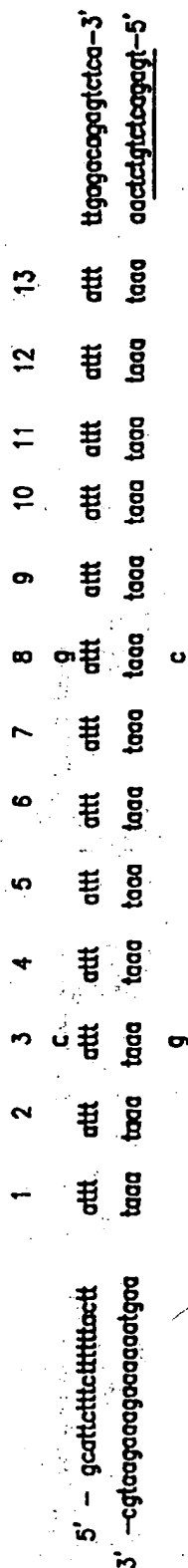
67 / 123

SEQUENCE OF THE AMPLIFIED 209 bp PCR-PRODUCT OF THE  $\beta$ -GLOBIN GENE

FORWARD PRIMER:  $\beta$ 2  
CATTGCTTC TGACAACT GTGTCTCA GCAACCTCA ACAGACACCA  
12mer PRIMER  
TGGTTCACCT GACTGCTGTG GAGAAGTCTG CCGTACTGC CCTGTGGGC  
AAGGTGAACG TGGATGAAGT TGGTGTGAG GCCCTGGCA GGTGCTATC  
AAGGTACAA GACAGGTTA AGGAGACCA TAGAACTGG GCATGTGGAG  
ACAGAGAAG  
REVERSE PRIMER  $\beta$ 11

**FIG. 51**

68/123



ALLELE NUMBER OF REPEATS	THEORETICALLY CALCULATED MOLECULAR MASS	
	ddG	ddC AND ddC
TRUNCATED	19440.60	11643.60
8xAAAT	15718.20	21033.60
9xAAAT	16959.00	22274.40
10xAAAT	18199.80	23515.20
11xAAAT	19440.60	24756.00
12xAAAT	20681.40	25996.80
13xAAAT	21922.20	27237.60

FIG. 52

69/123

FIG. 53A

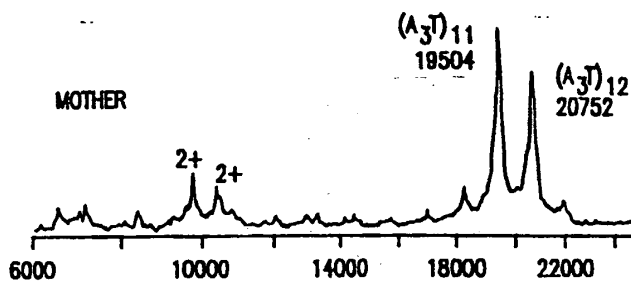


FIG. 53B

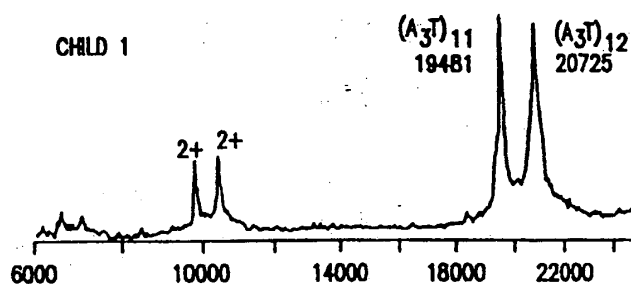


FIG. 53C

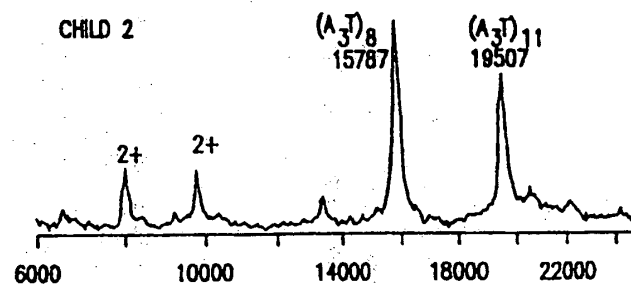


FIG. 53D

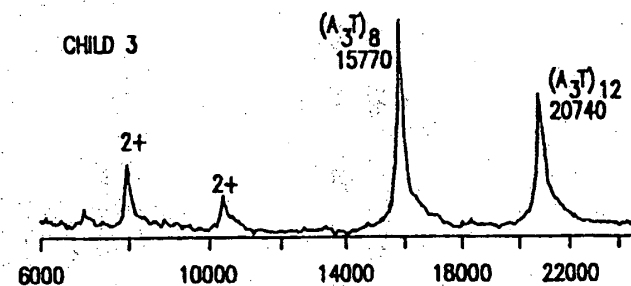
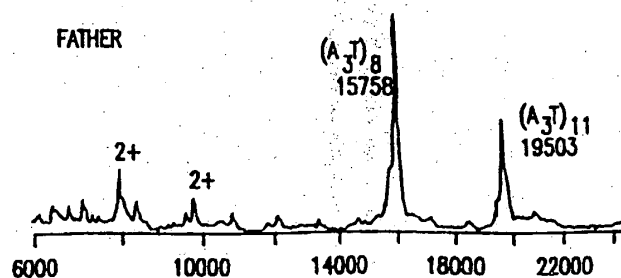
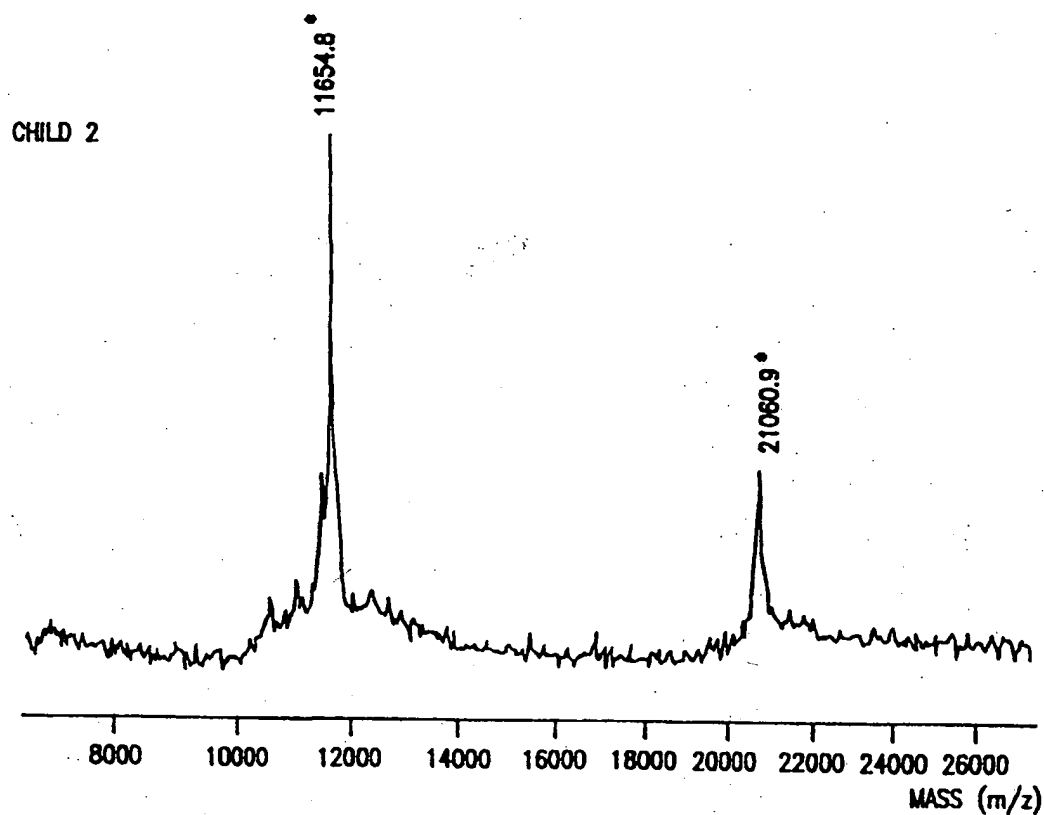
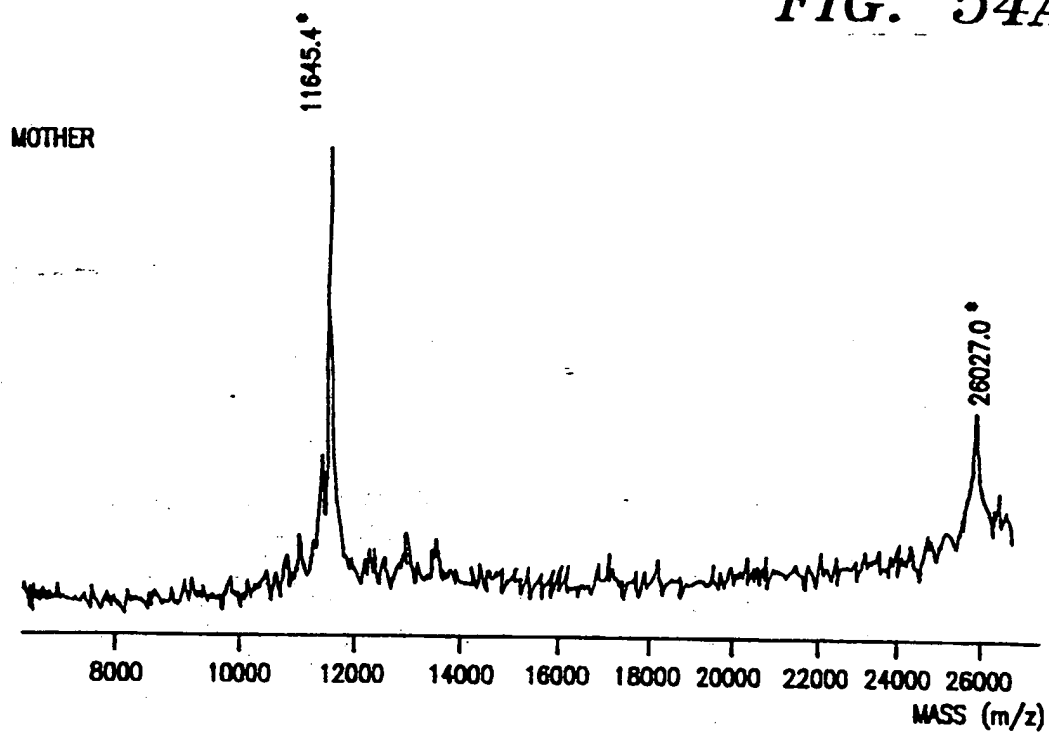


FIG. 53E



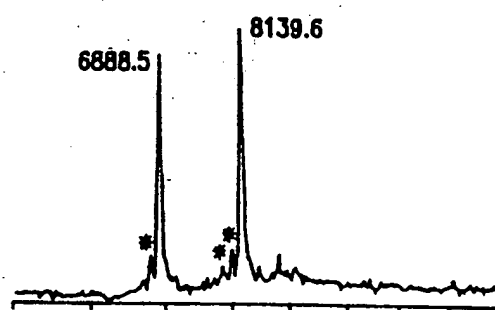
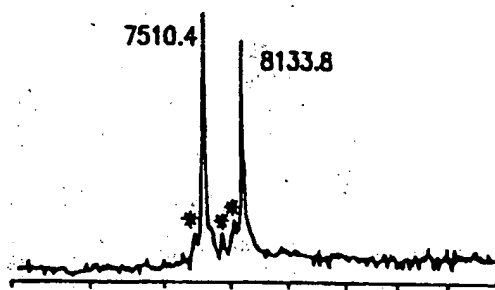
70/123

*FIG. 54A*

71/123

5' -GTGTGTGTGTGTGTGTGTTTT (TT) (TT) AACAGGGATTGGGAATTATTGAGA-3'  
PRIMER TTGTCCTAAACCCCTT (4448.0)  
T5 ALLELE CAAAA --- TTGTCCTAAACCCCTT (6890.6)  
T7 ALLELE CAAAA AA --- TTGTCCTAAACCCCTT (7515.0)  
T9 ALLELE CAAAA AA AA TTGTCCTAAACCCCTT (8139.4)

*FIG. 55*

*FIG. 56A**FIG. 56B*

73/123

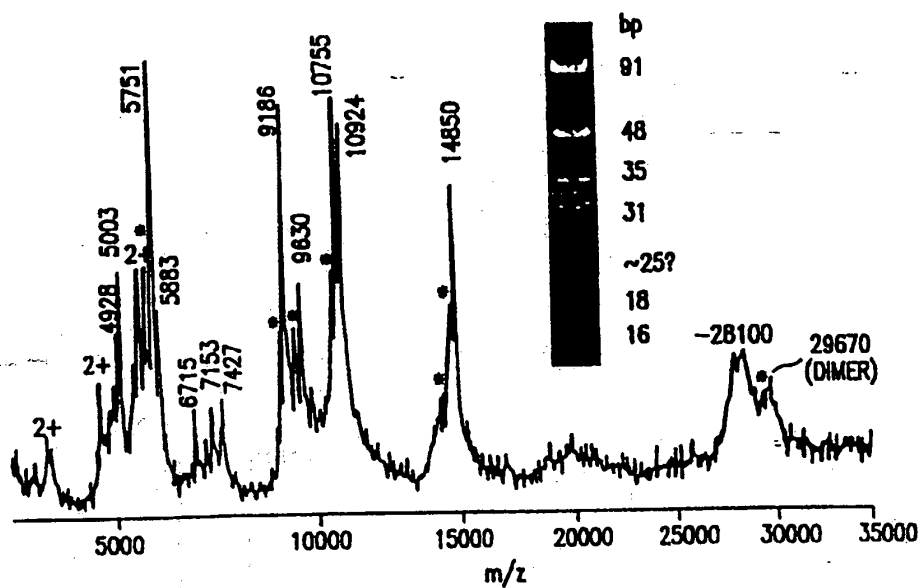


FIG. 57A

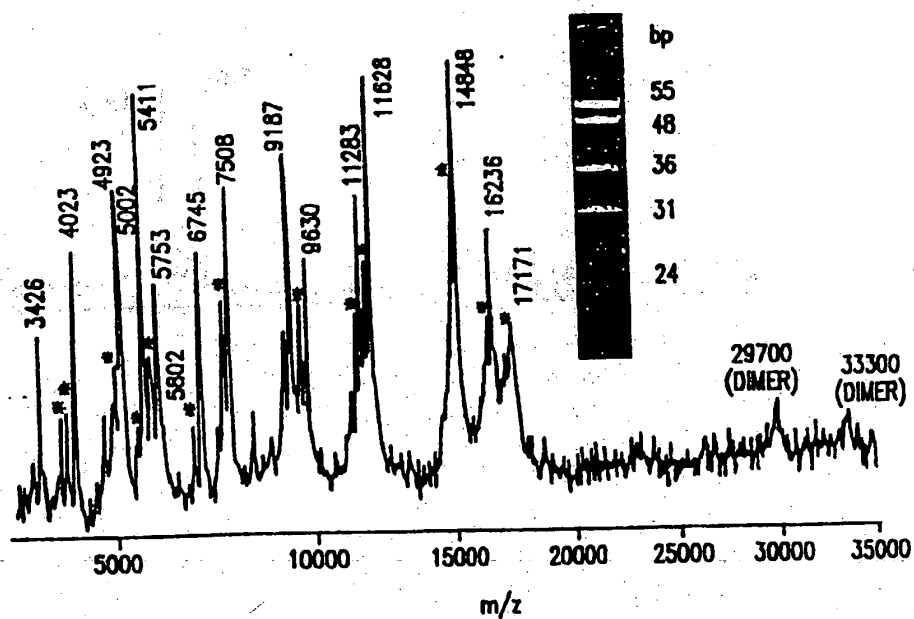


FIG. 57B



74 / 123

FIG. 58A

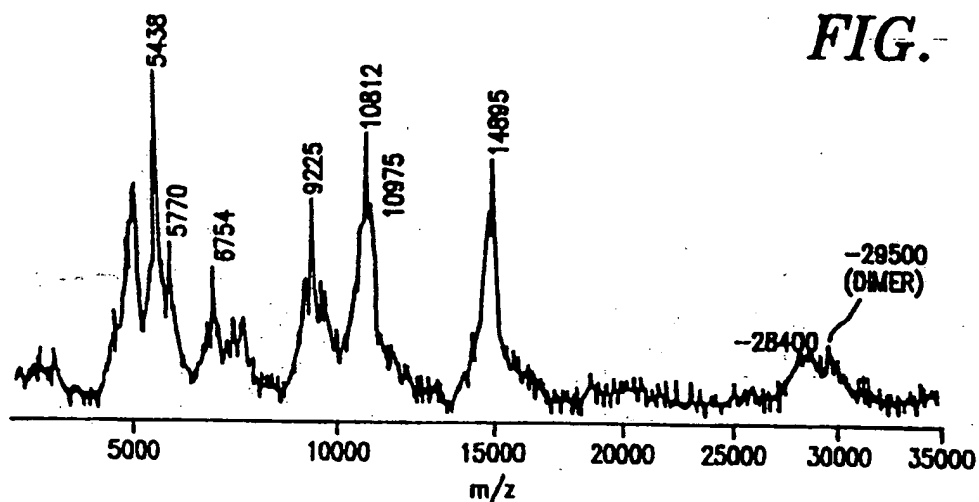


FIG. 58B

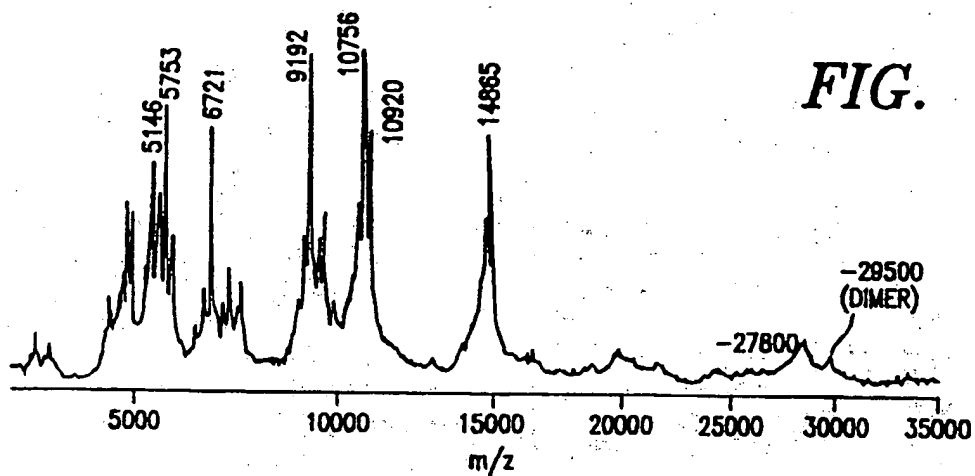
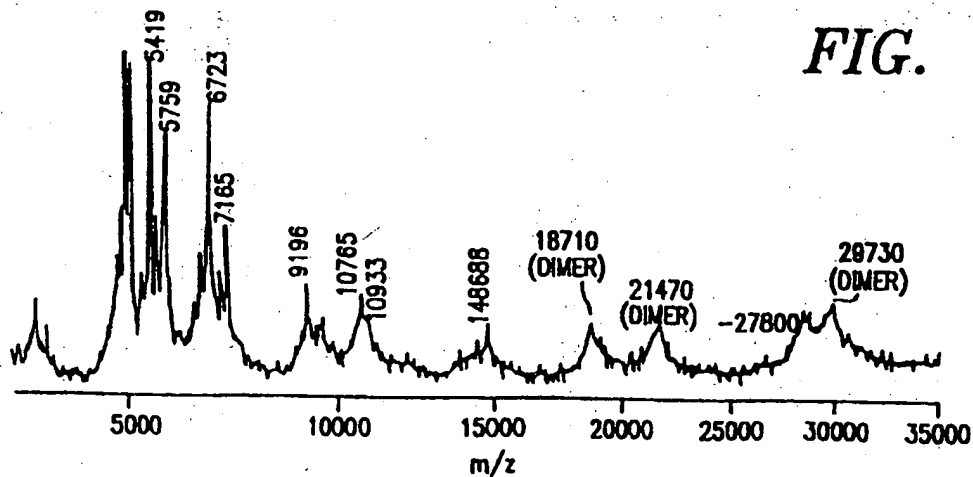


FIG. 58C



75/123

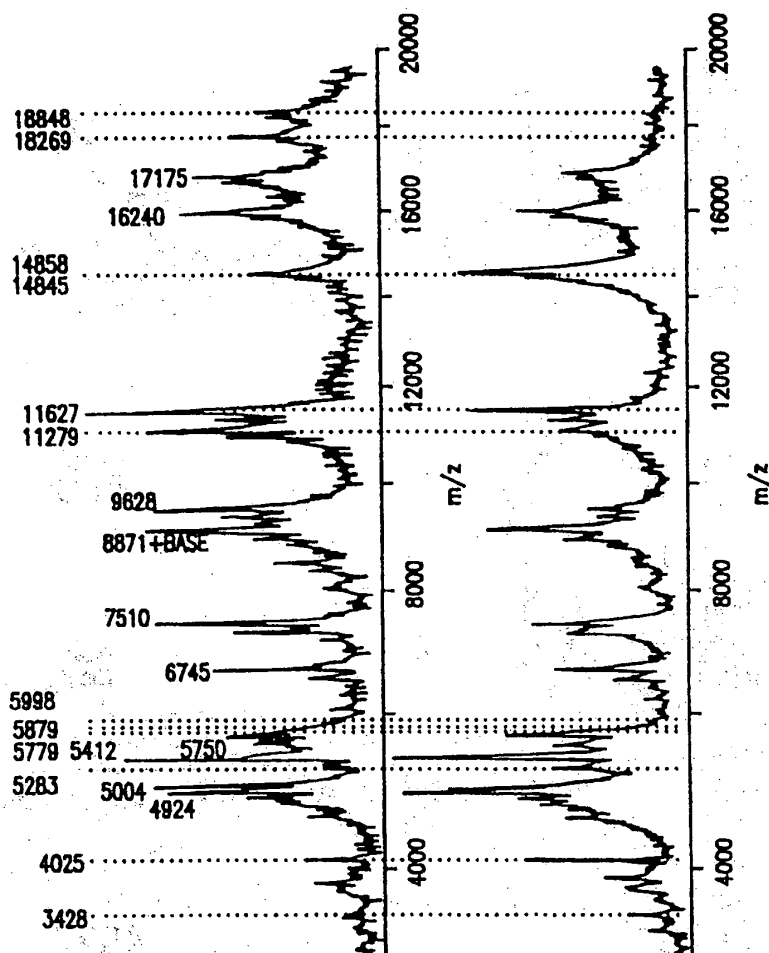


FIG. 59A

FIG. 59B

76/123

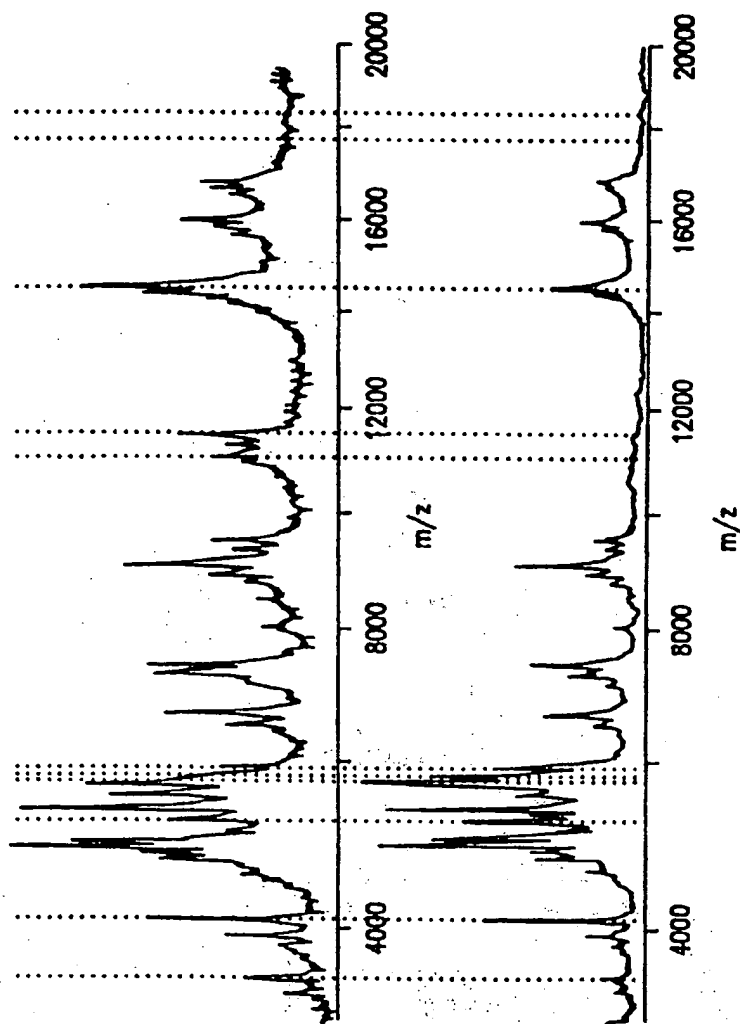


FIG. 59C

FIG. 59D

77/123

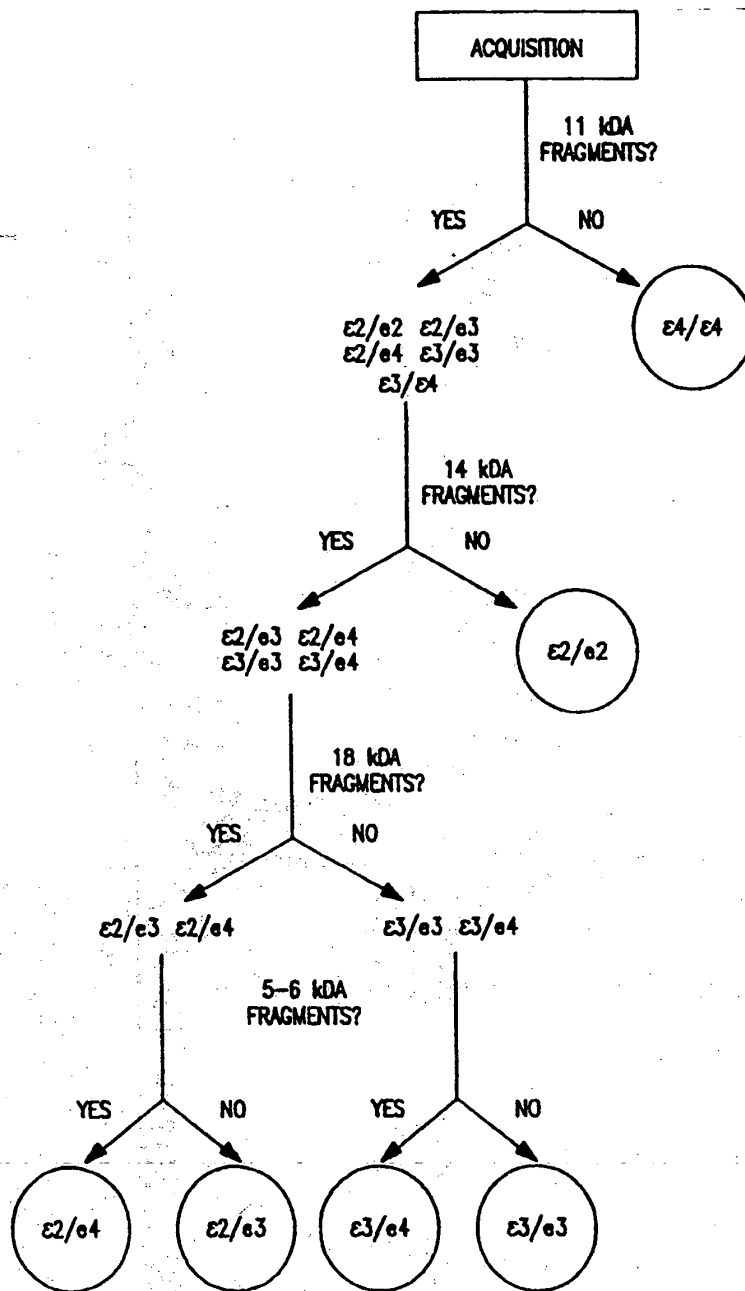


FIG. 60



79/123

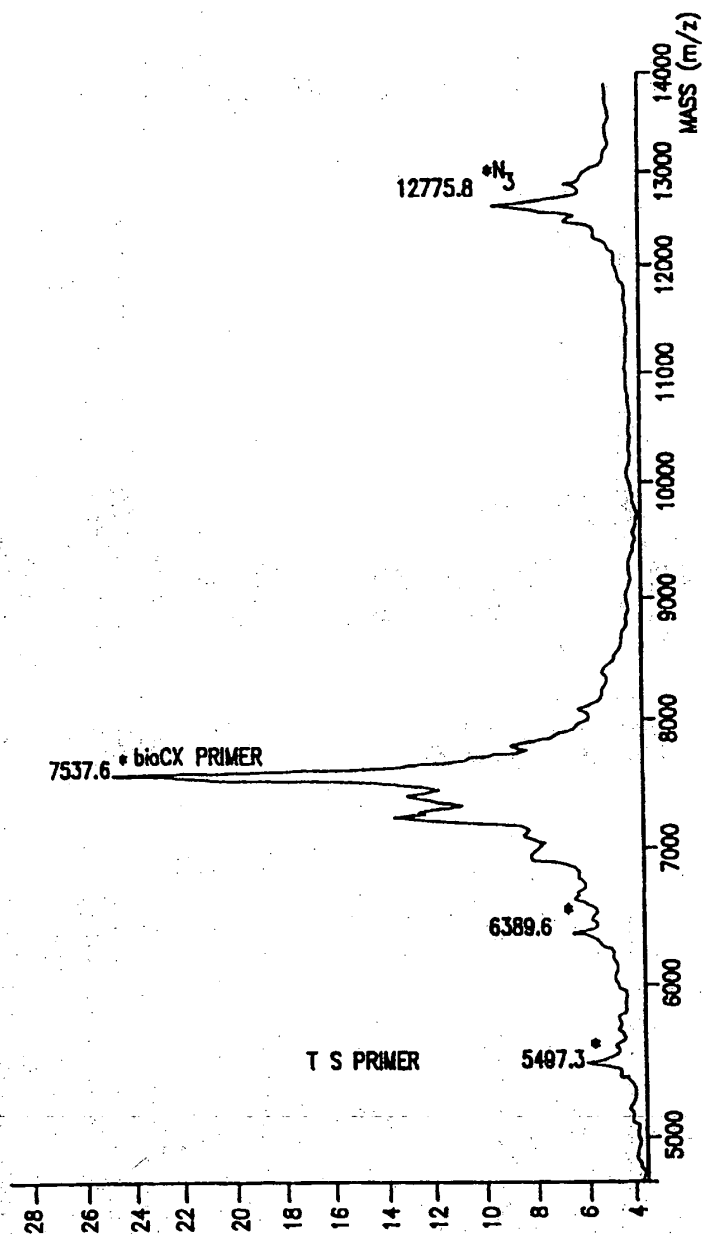


FIG. 62

80/123

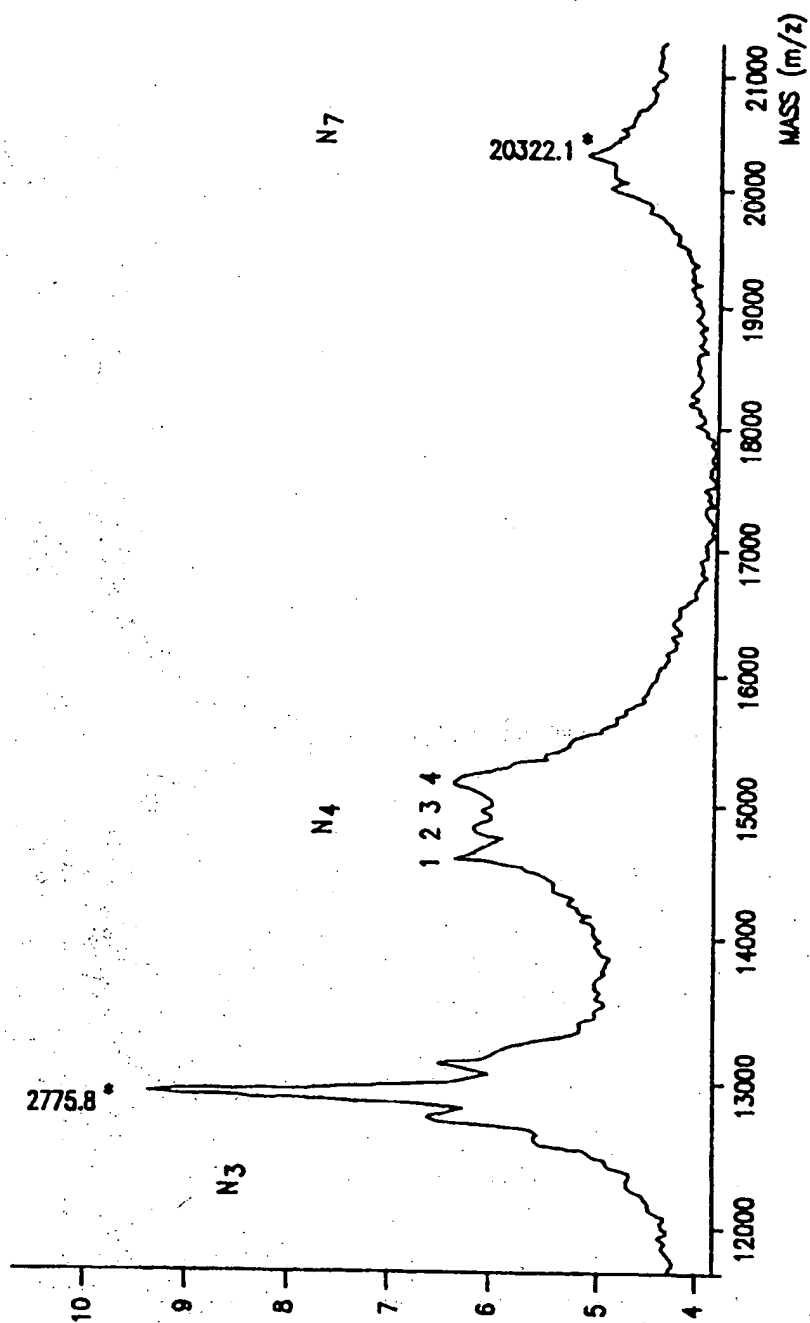


FIG. 63

81/123

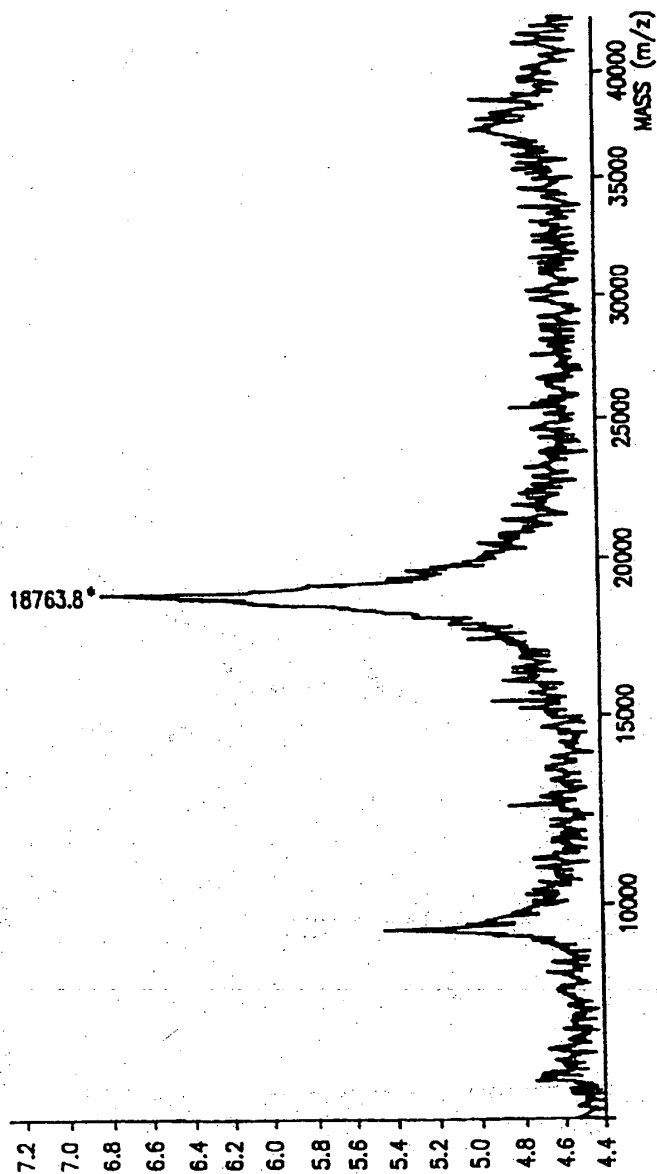


FIG. 64



82/123

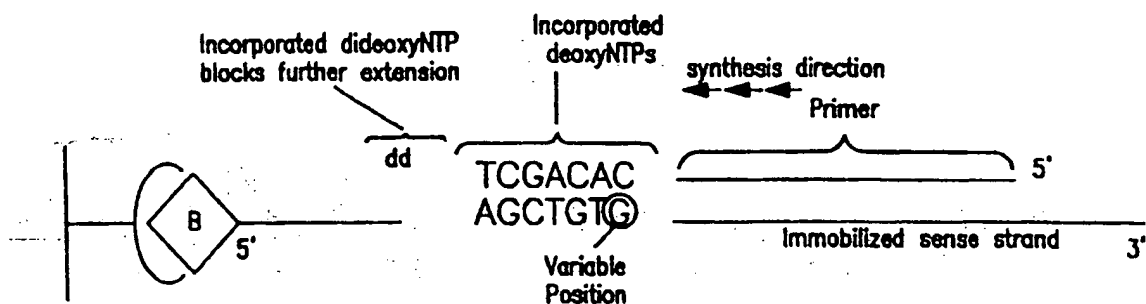


FIG. 65A

<u>ddTTP + dNTP (N=A,C,G)</u>		<u>ddATP + dNTP (N=C,T,G)</u>
<u>CGG CTG CGA TCA CCG TGC GG</u> C ACA GCT		<u>CGG CTG CGA TCA CCG TGC GG</u> C A
WILDTYPE 8246 Da		WILDTYPE 6721 Da
<u>CGG CTG CGA TCA CCG TGC GG</u> T		<u>CGG CTG CGA TCA CCG TGC GG</u> T A
8423 Da		6736 Da
<u>CGG CTG CGA TCA CCG TGC GG</u> A ACA GCT		<u>CGG CTG CGA TCA CCG TGC GG</u> A
8270 Da		6432 Da

FIG. 65B

83/123

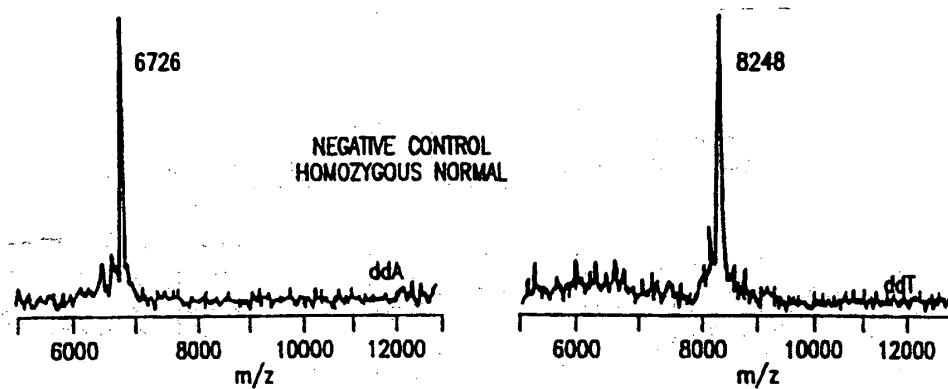


FIG. 66A

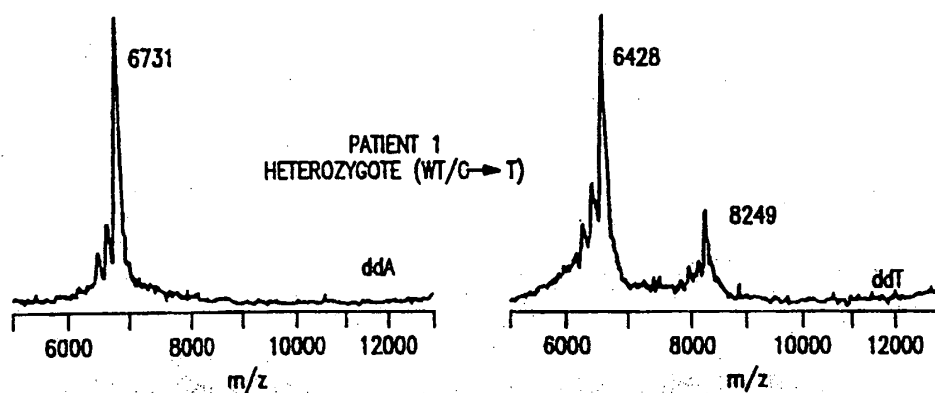


FIG. 66B

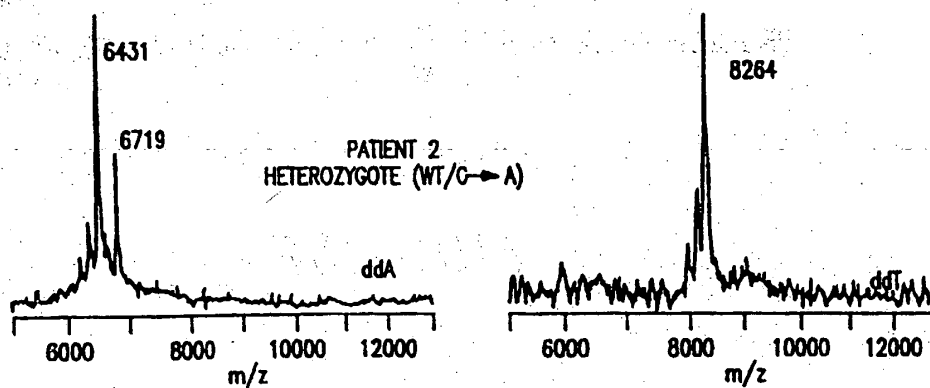
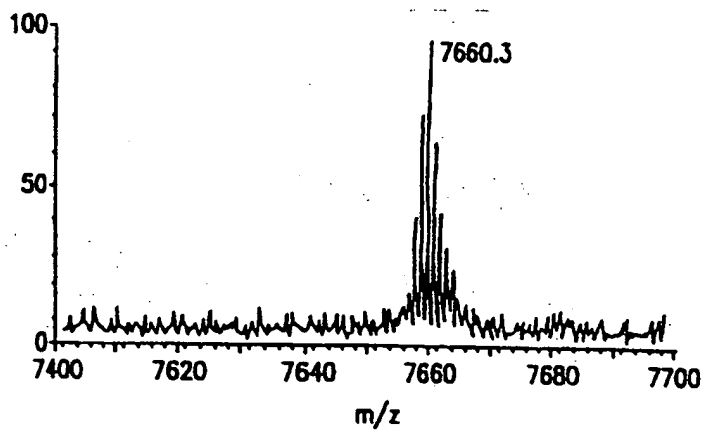
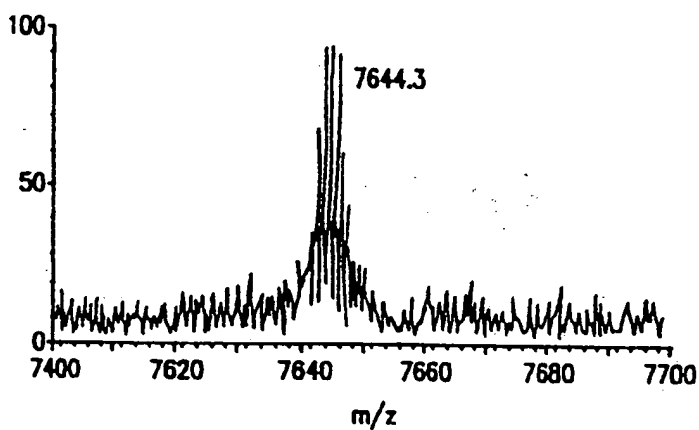
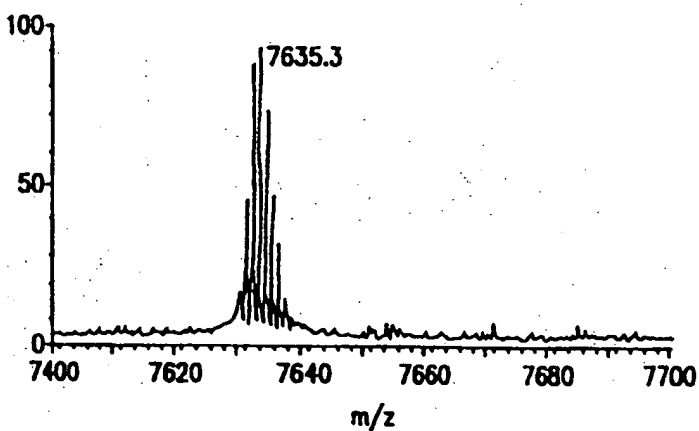
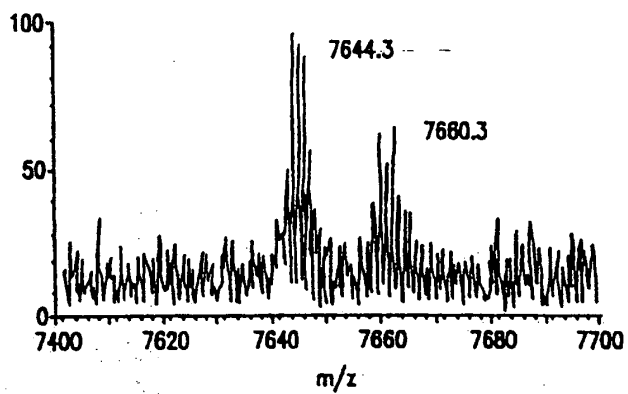
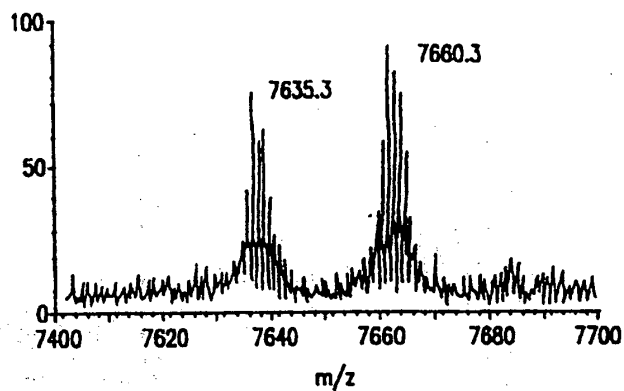
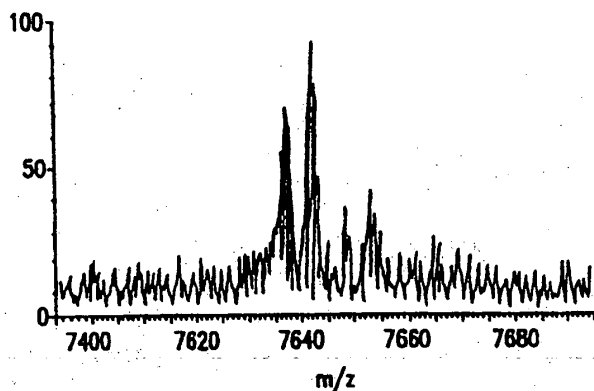


FIG. 66C

*FIG. 67A**FIG. 67B**FIG. 67C*

85/123

*FIG. 67D**FIG. 67E**FIG. 67F*

86 / 123

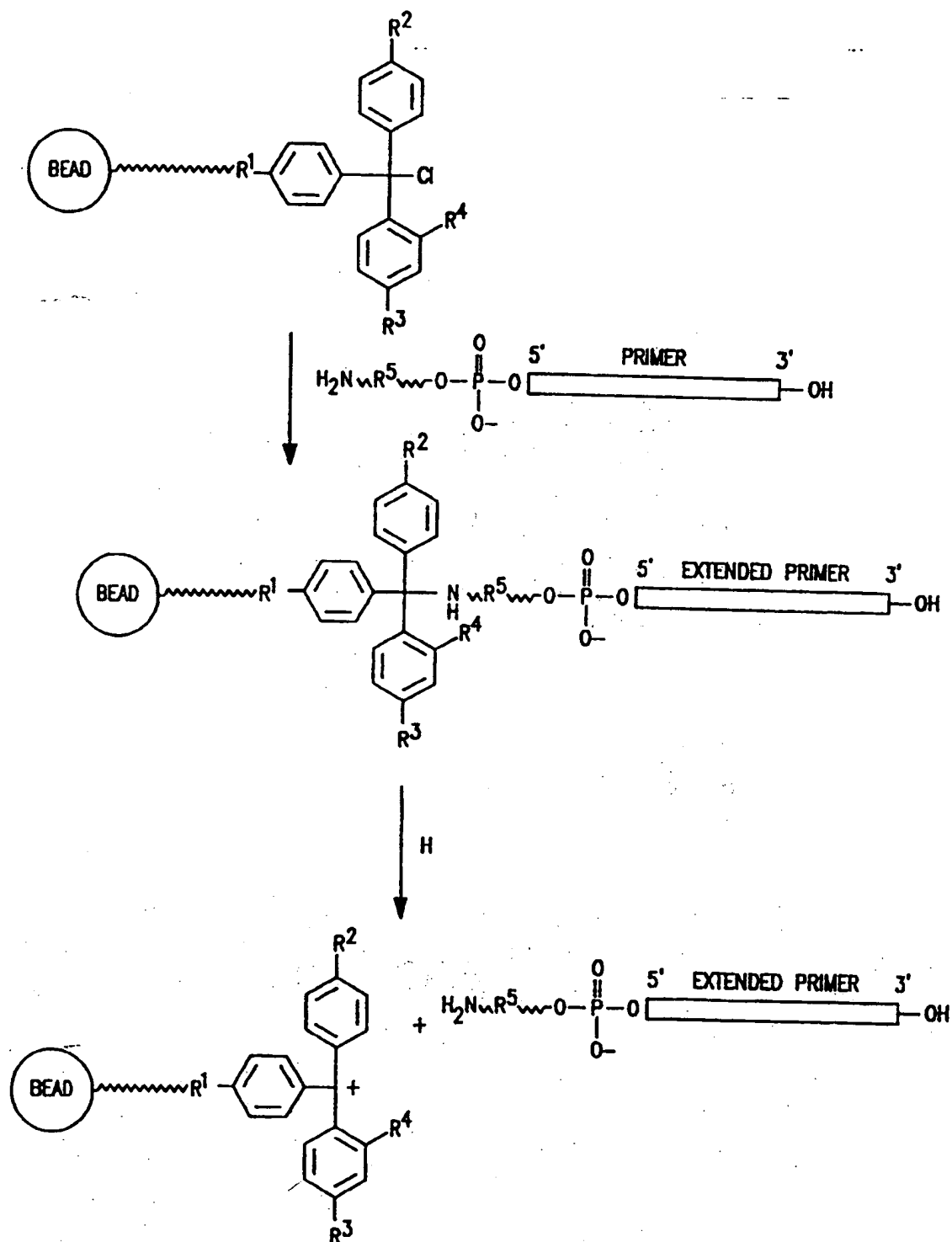
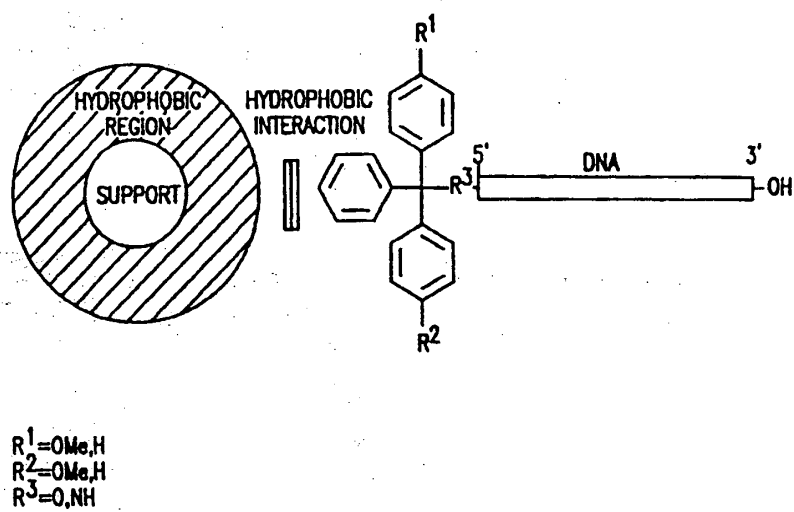
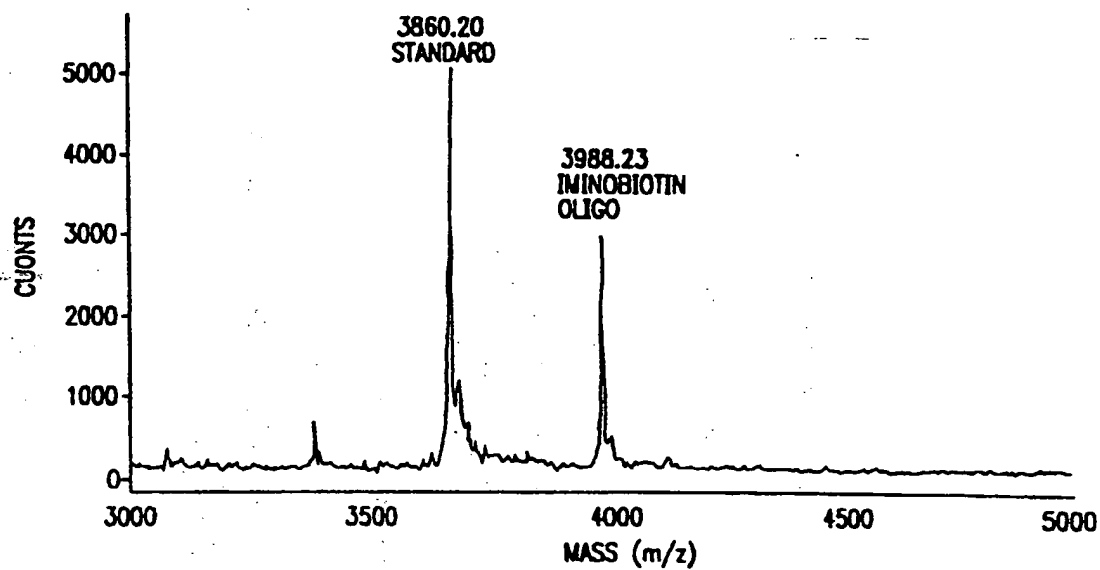
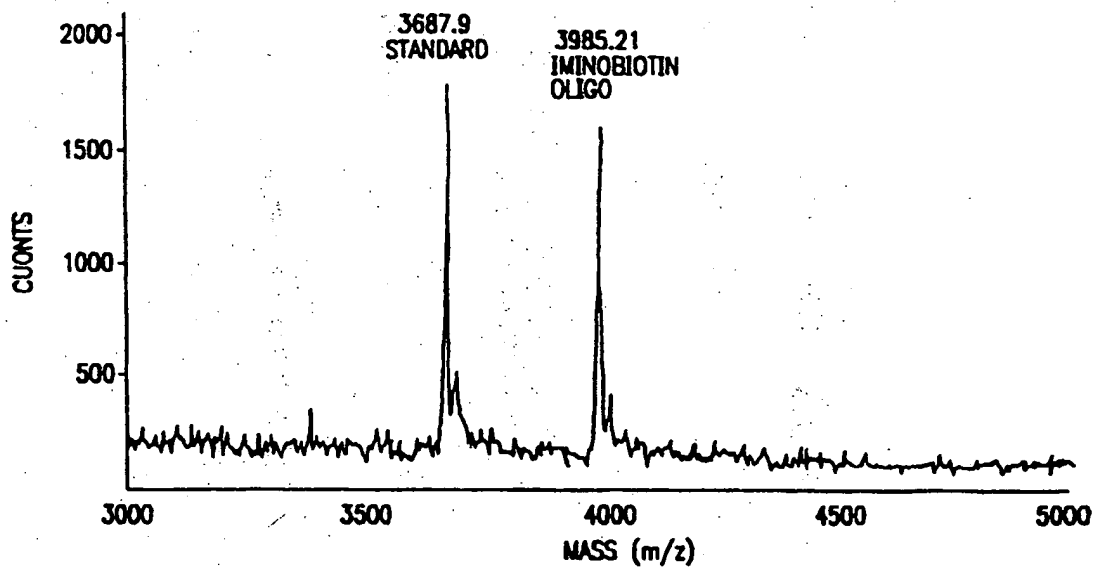


FIG. 68

87/123

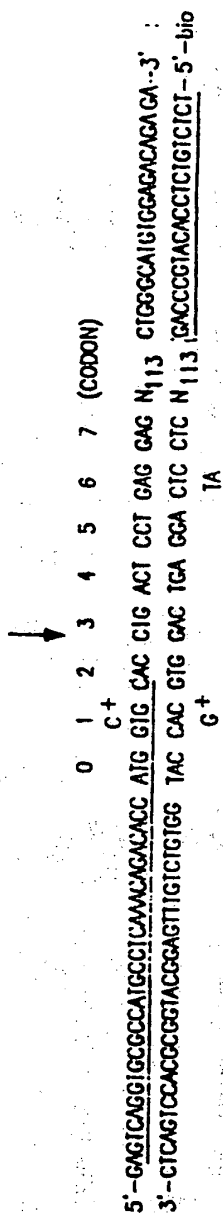
**FIG. 69**

88/123

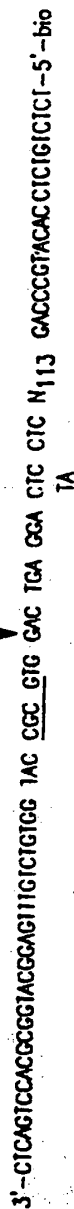
*FIG. 70**FIG. 71*

89/123

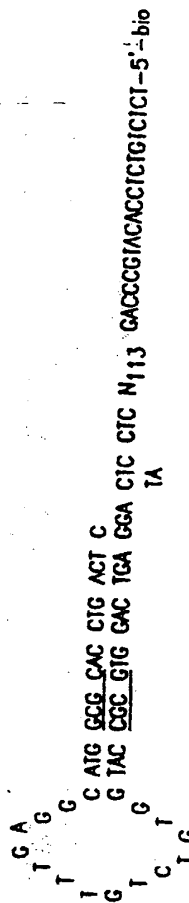
## FIG. 72A

LOOP-PROBE FOR MUTATION DETECTION IN THE  $\beta$ -GLOBIN GENE AT CODON 5 AND CODON 6a) PCR-AMPLIFICATION USING  $\beta$ -GLOBIN GENE SPECIFIC PRIMERS WITH MODIFICATIONS

b) DENATURATION OF BIOTIN-STREPTAVIDIN CAPTURED dsDNA



c) ANNEALING OF COMPLEMENTARY 3'-END (LOOP)







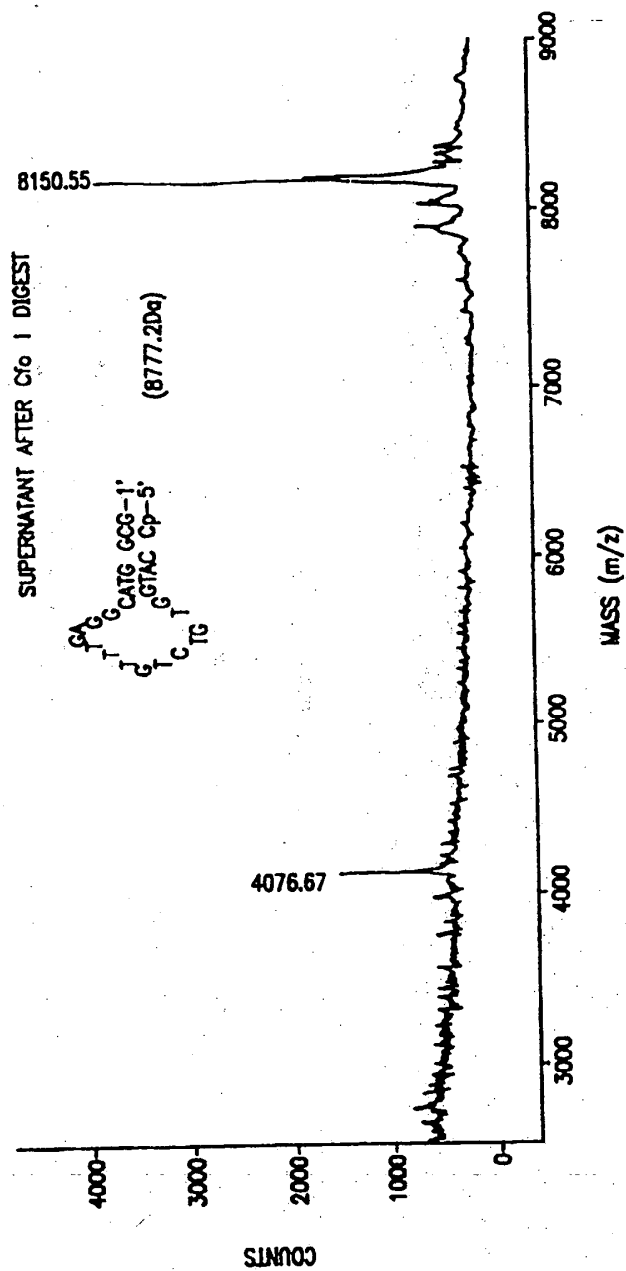


FIG. 73A

92/123

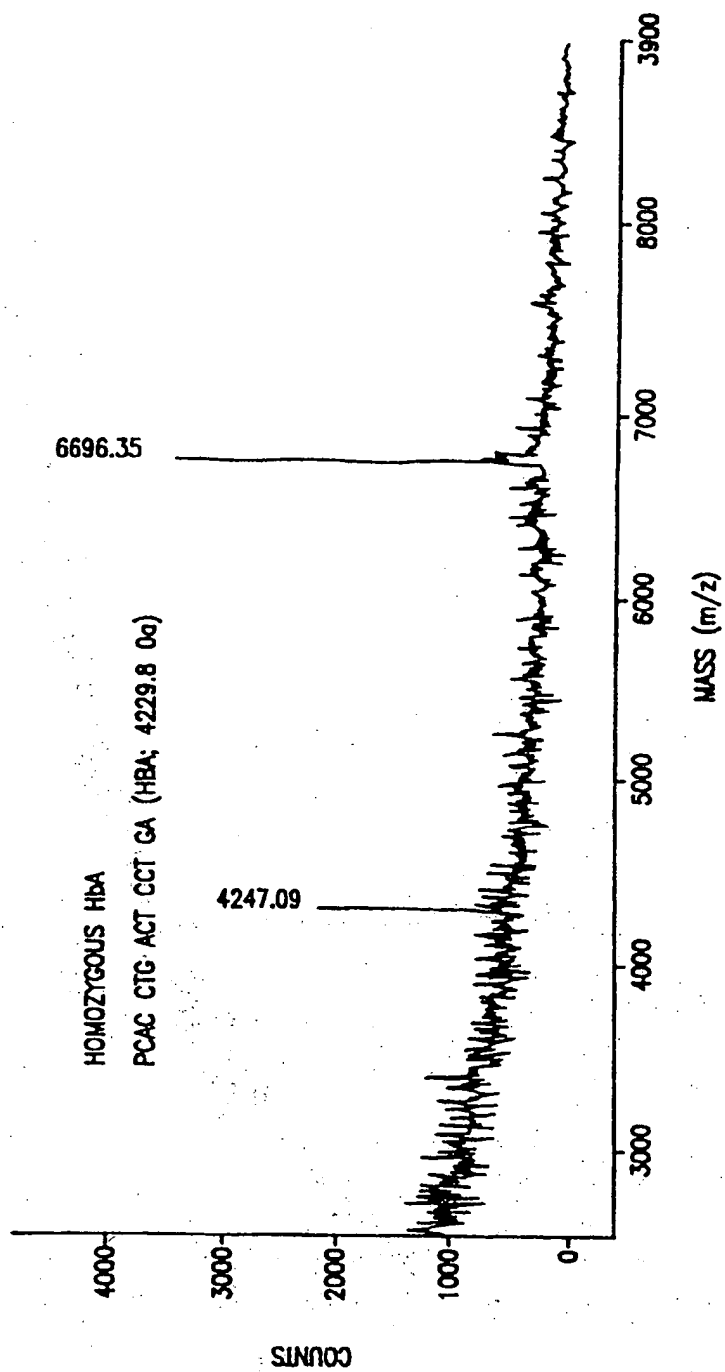


FIG. 73B

93/123

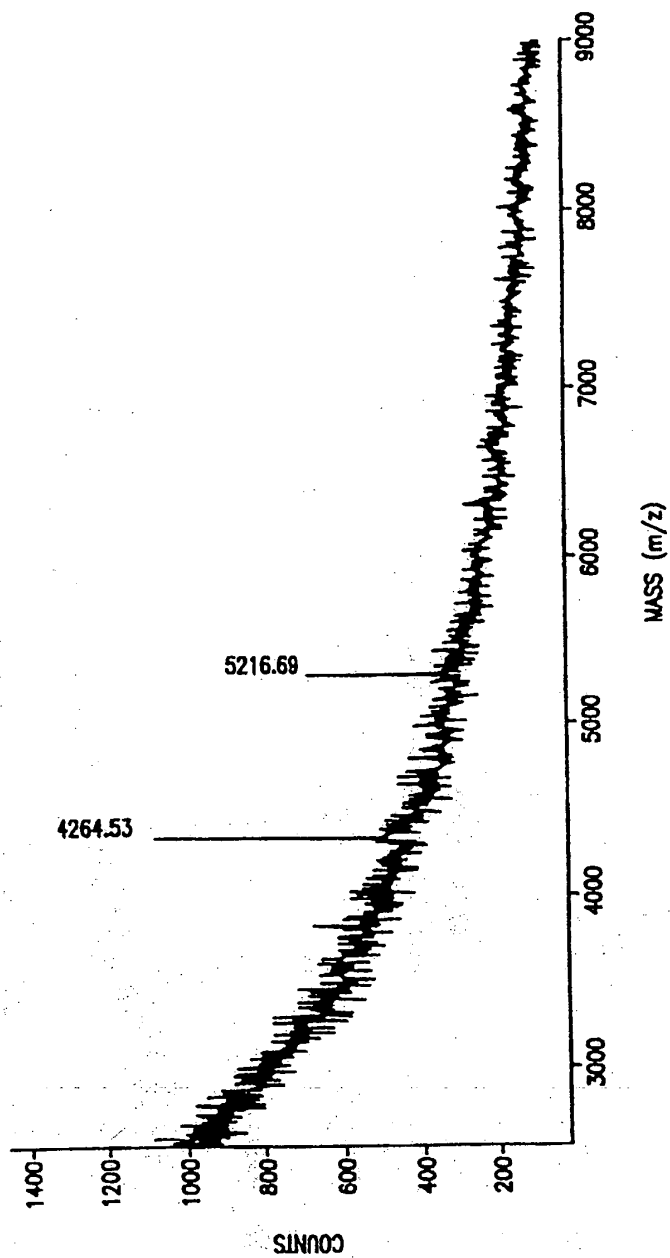


FIG. 73C

94/123

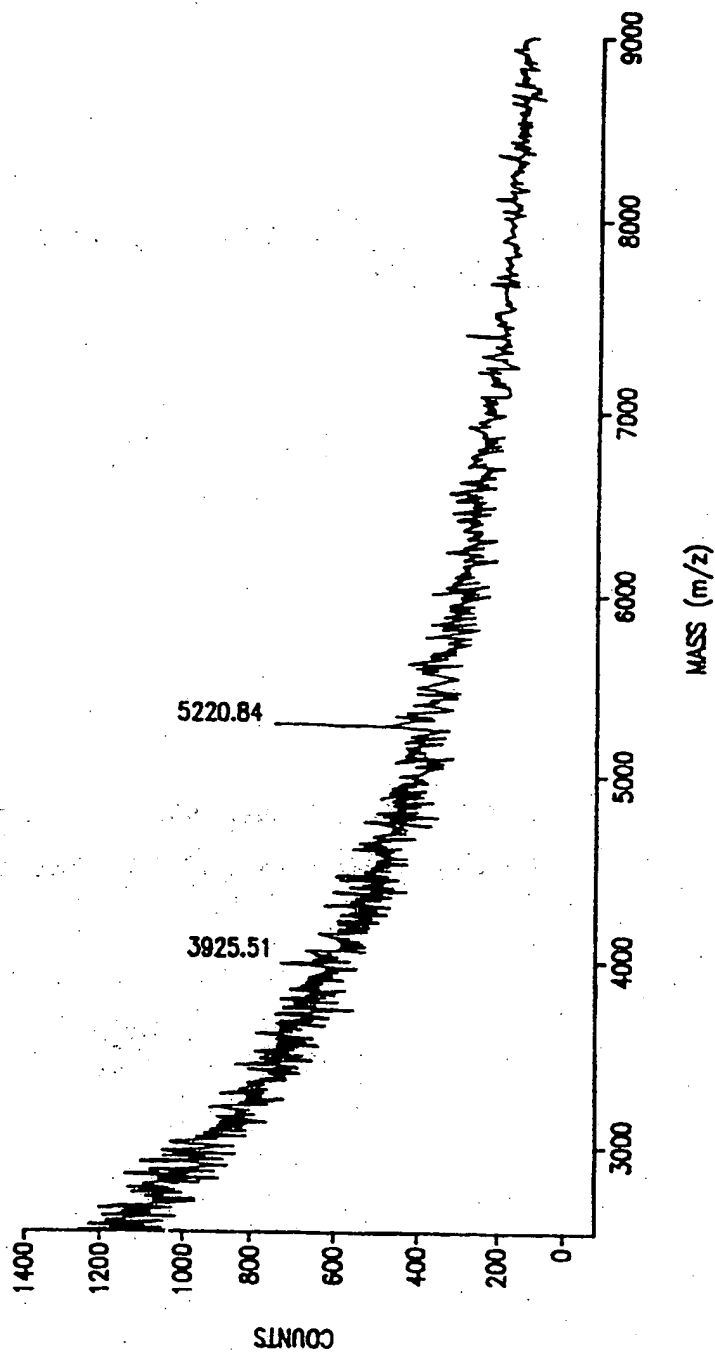


FIG. 73D

95/123

5' CAGCTCTCATTTCCATACAGTCAGTATCAATTCTGGAAGAATTCCAGACATTAA  
GATAGTCATCTGGGGCT 3'

*FIG. 74*

5' ACCTACCGTTCAGTTCGACTGAGATAATACGACTCATATAGCAGCTCTCATTTCCATAC3'  
RANDOM SEQUENCE T7 PROMOTER SEQUENCE CKR-5 HOMOLOGUE

*FIG. 75*

96/123

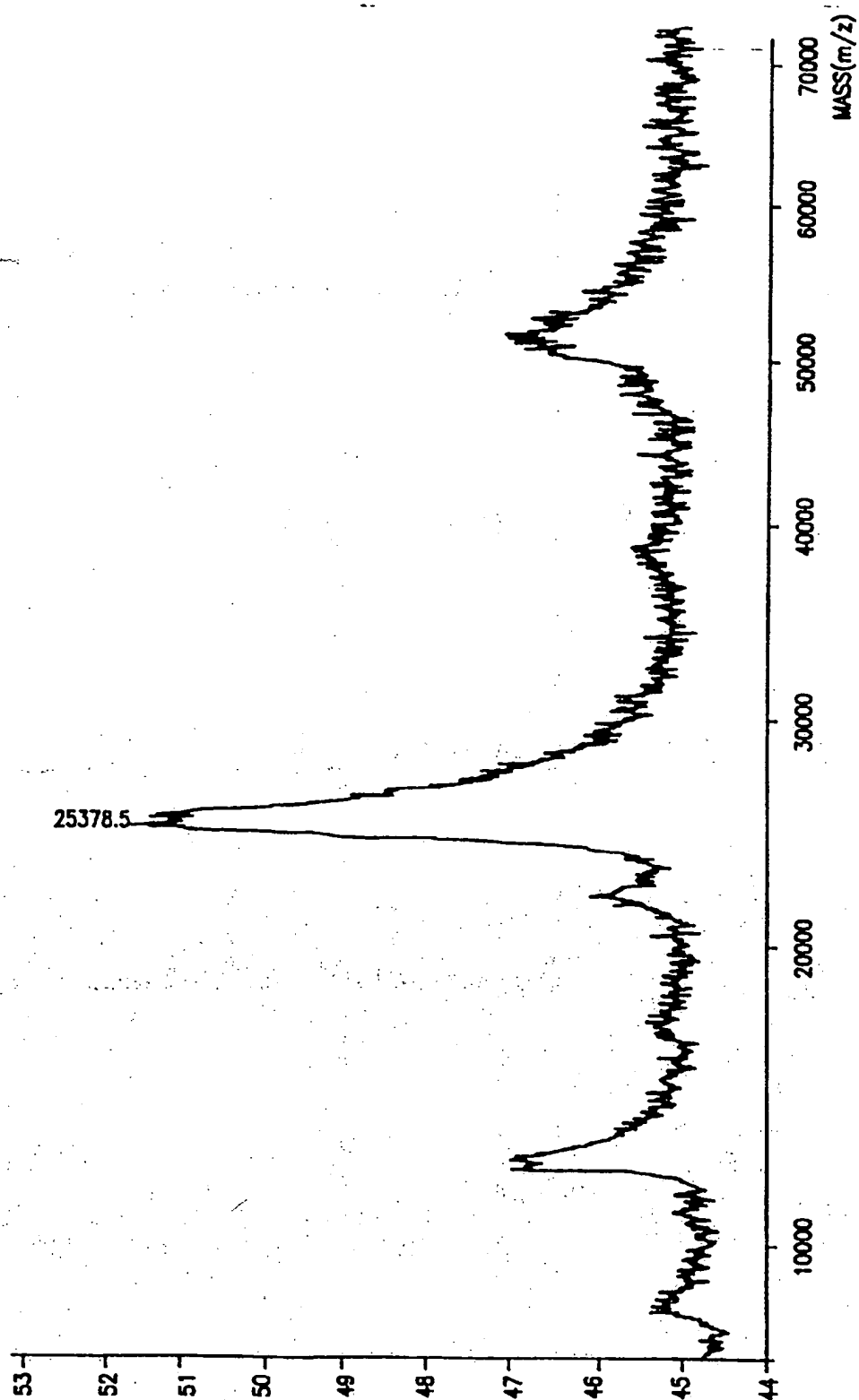
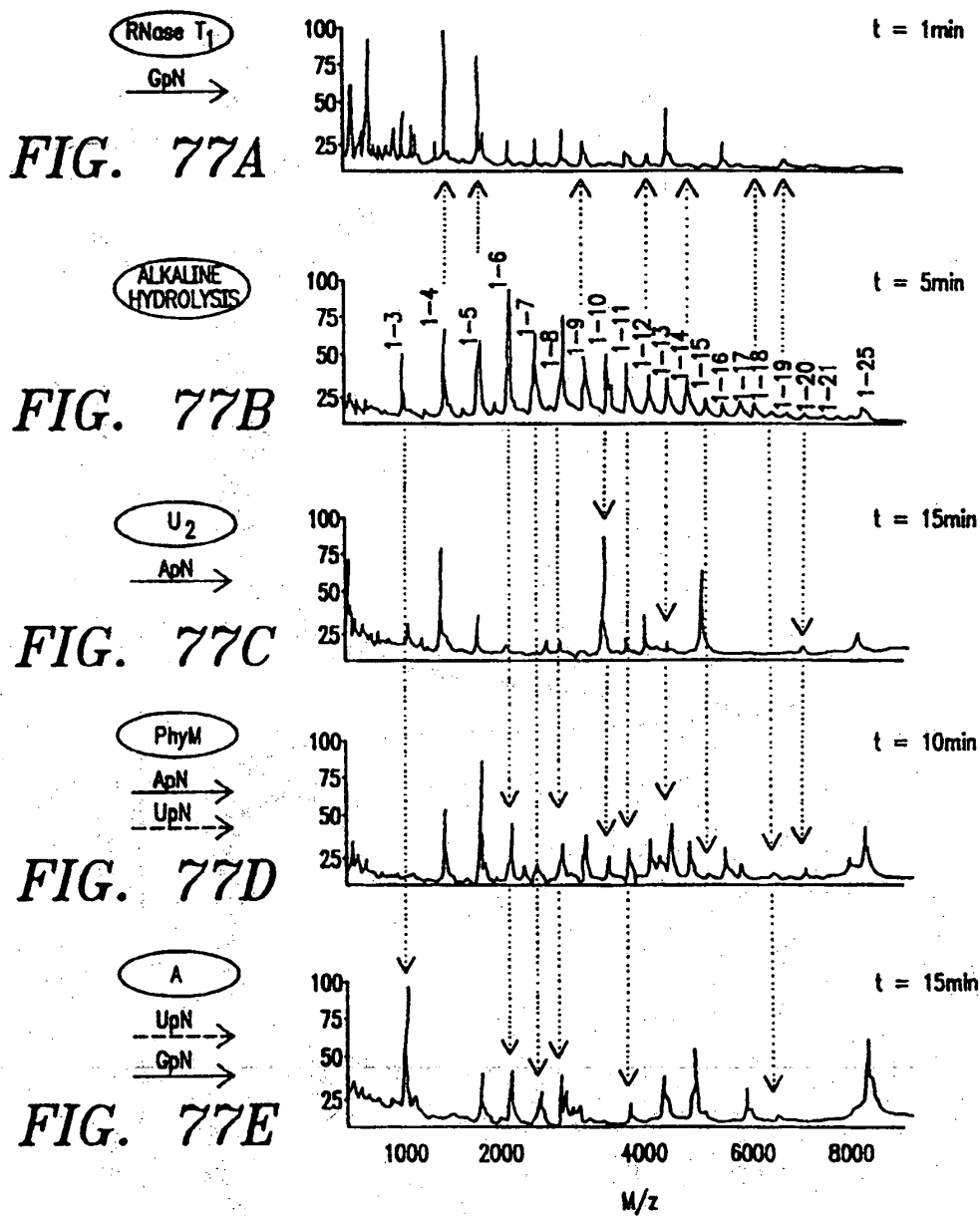


FIG. 76

97/123





98/123

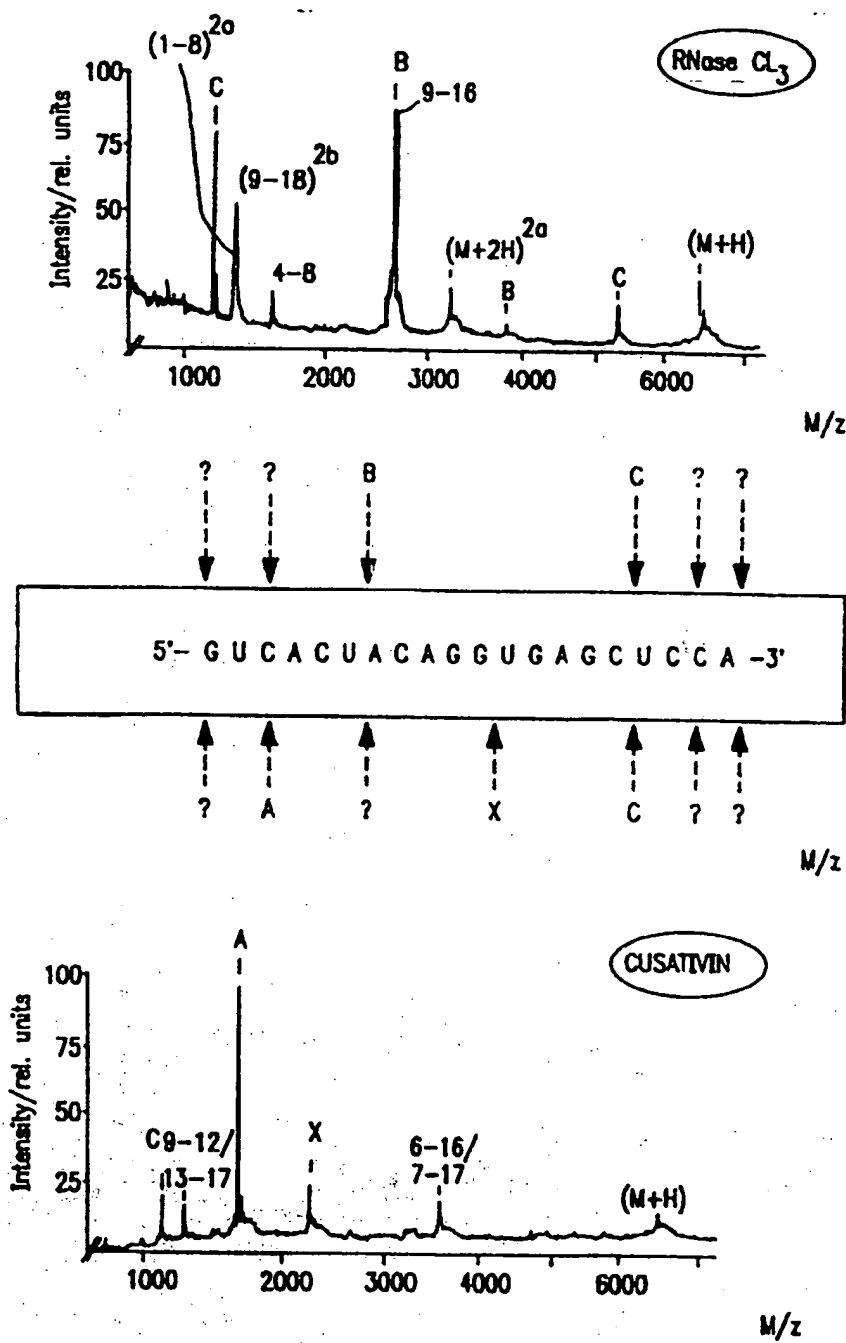


FIG. 78

99/123

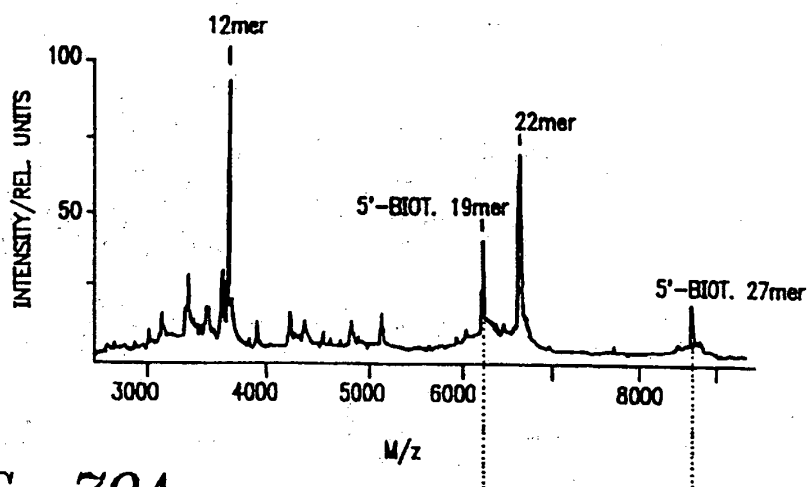


FIG. 79A

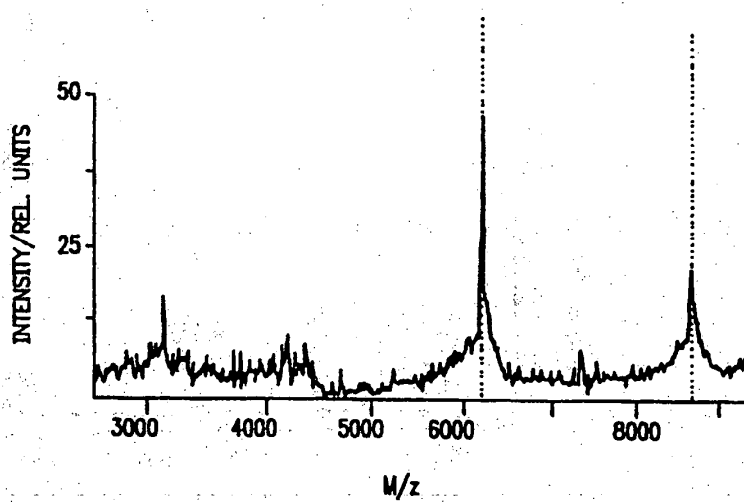
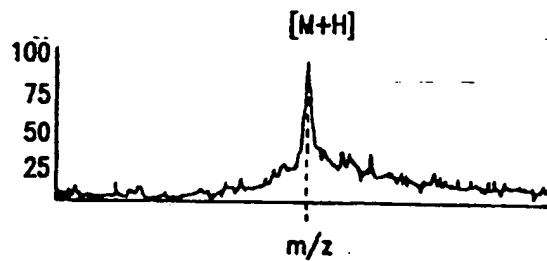
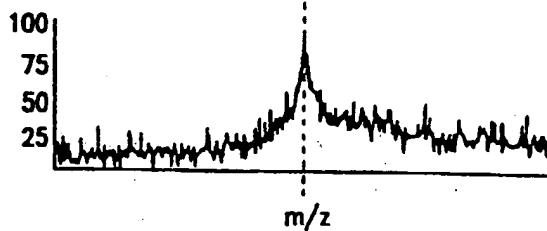
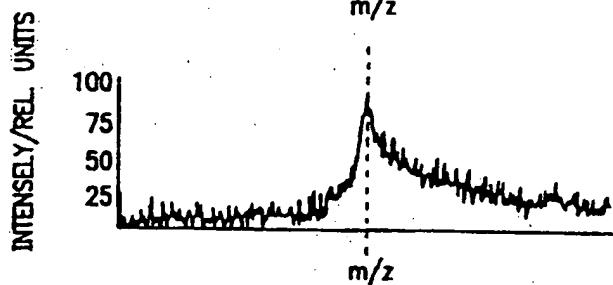
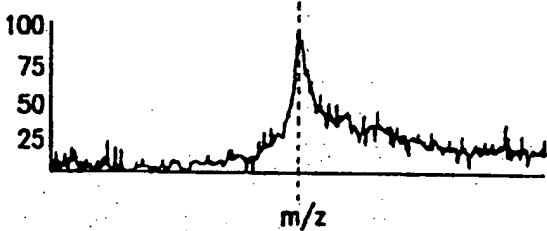
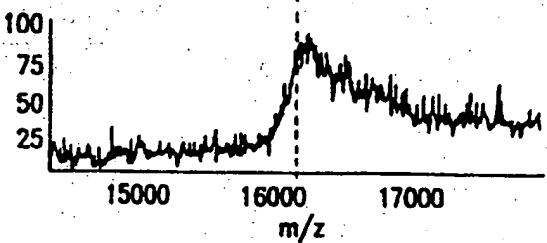


FIG. 79B

100/123

*FIG. 80A**FIG. 80B**FIG. 80C**FIG. 80D**FIG. 80E*

101/123

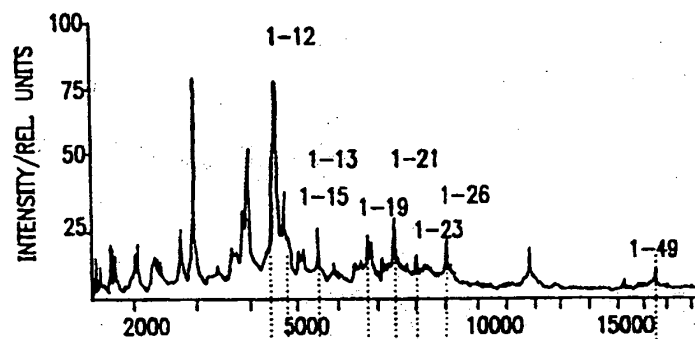


FIG. 81A

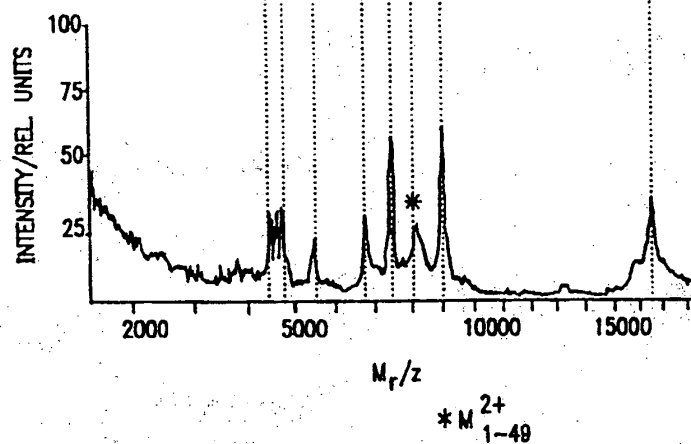


FIG. 81B

FIG. 82A

$\beta$ -2  
CATTGCTTC TCACACACT GTGTTCACTA GCACCTCA ACACACACC ATG GTG CAC CTG ACT C  
 $\beta$ -TAG1  
 GTGTCGC ATG GTG CAC CTG ACT C (CODON)  
 CCT GAG GAG AG TCT GCC GTT  
 THALASSEMIA (COD5 ACT)  
 A HbC ---  
 T HbS  
  
 $\beta$ -TAG2  
 CGCTGTGT GAG GCC CTG GGC A (CODON)  
 ACT GCC CTG TGG GGC AAG GTG AAC GTG GAT GAA GTT GGT GAG GCC CTG GGC AG GTTGTATCA AGGTACAG  
 A THALASSEMIA (IVS -1 nt1 G>A)  
 C THALASSEMIA (IVS -1 nt5 G>C)  
 CTHALASSEMIA (IVS -1 nt6 T>C)

ACAGGTTAA GGACACCAAT AGAACTGGG CATGTGGAGA CAGAGAG  
 $\beta$  -11

FIG. 82B

(1) SINGLE STRANDED  
 (STV) BIO3' - ACCTCAGTCCATCGTCAGCAGCAG - 5'  
 UNIVERSAL SS SEQUENCE | COMPLEMENTARY  
 E.G.  $\beta$ -TAG1  
  
 (2) PARTLY DOUBLE STRANDED  
 5'-TGCAGTCAGGTACAGTC-3'  
 (STV) BIO 3'-ACCTCAGTCCATCGTCAGCAGCAG - 5'  
 UNIVERSAL DS SEQUENCE | COMPLEMENTARY  
 E.G.  $\beta$ -TAG2

103/123

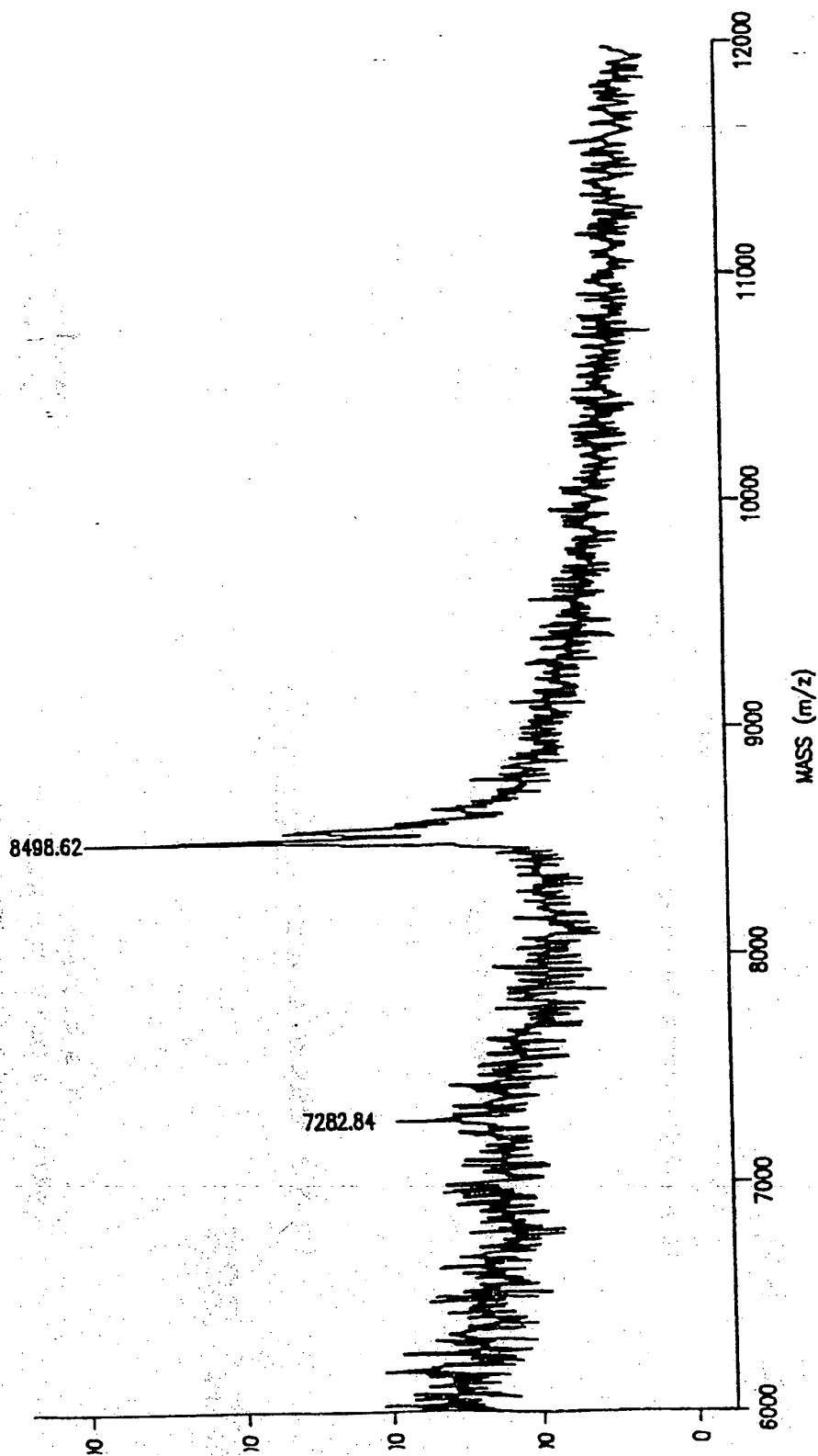


FIG. 83

104/123

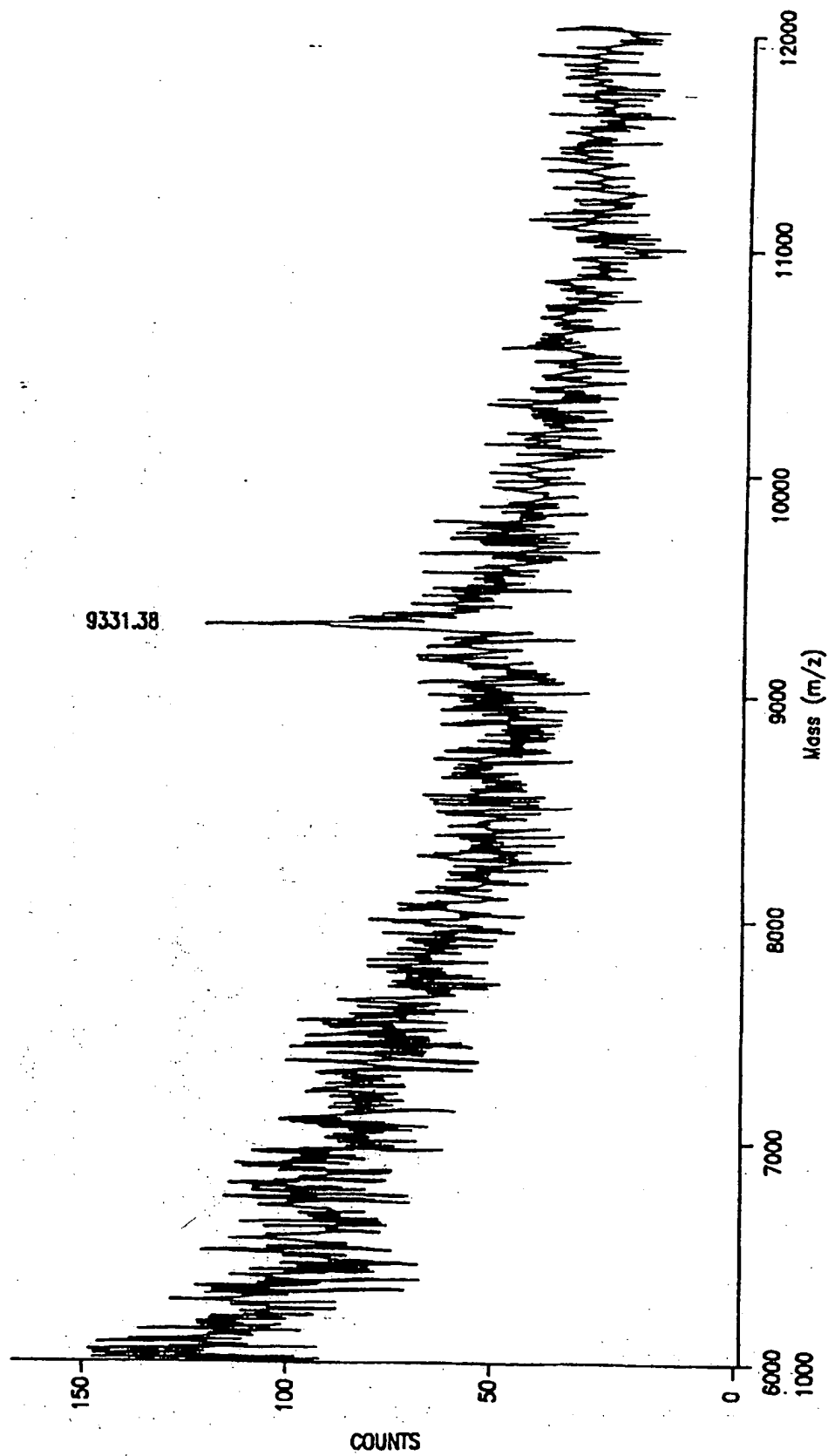


FIG. 84

105/123

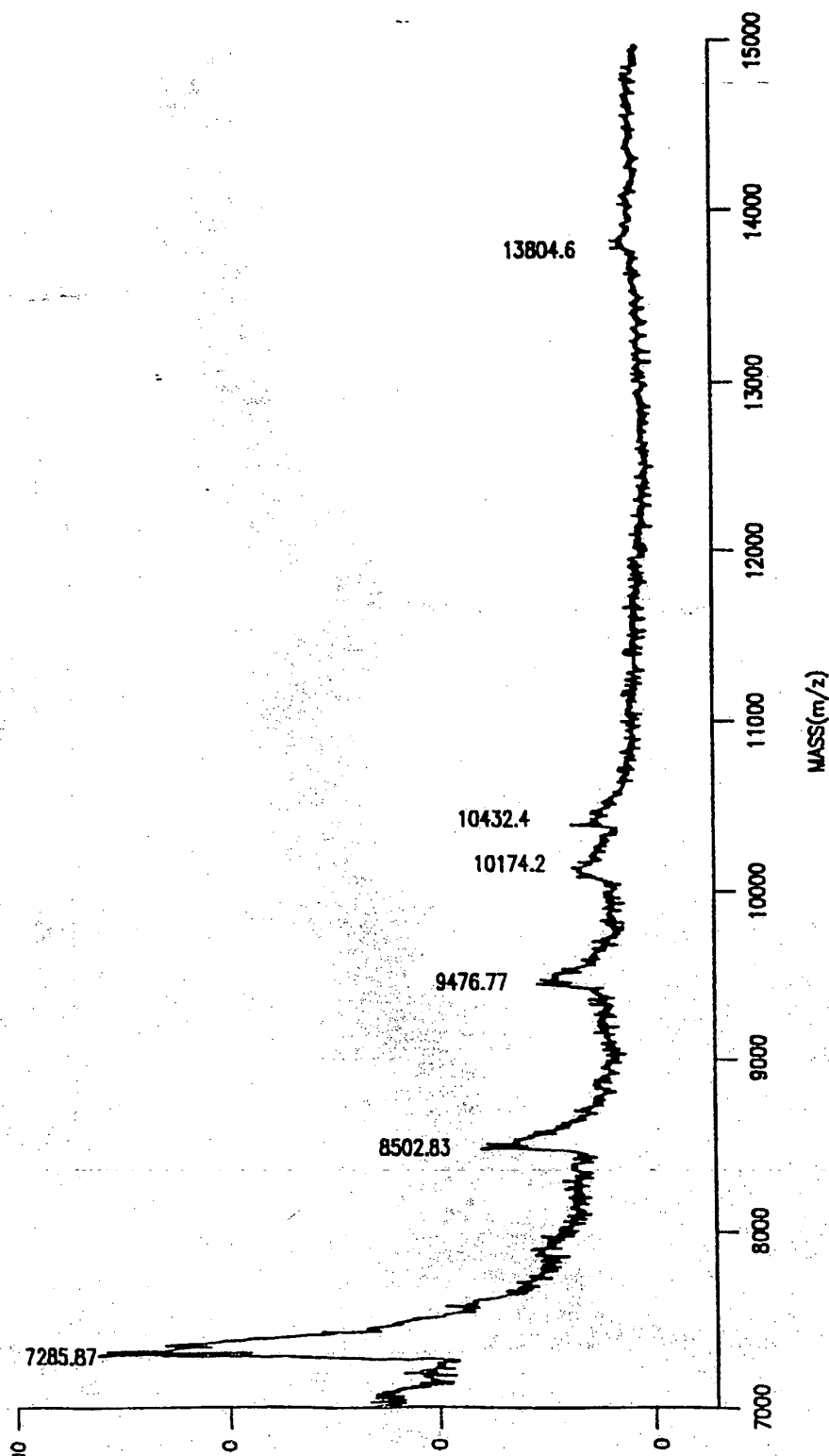


FIG. 85



106/123

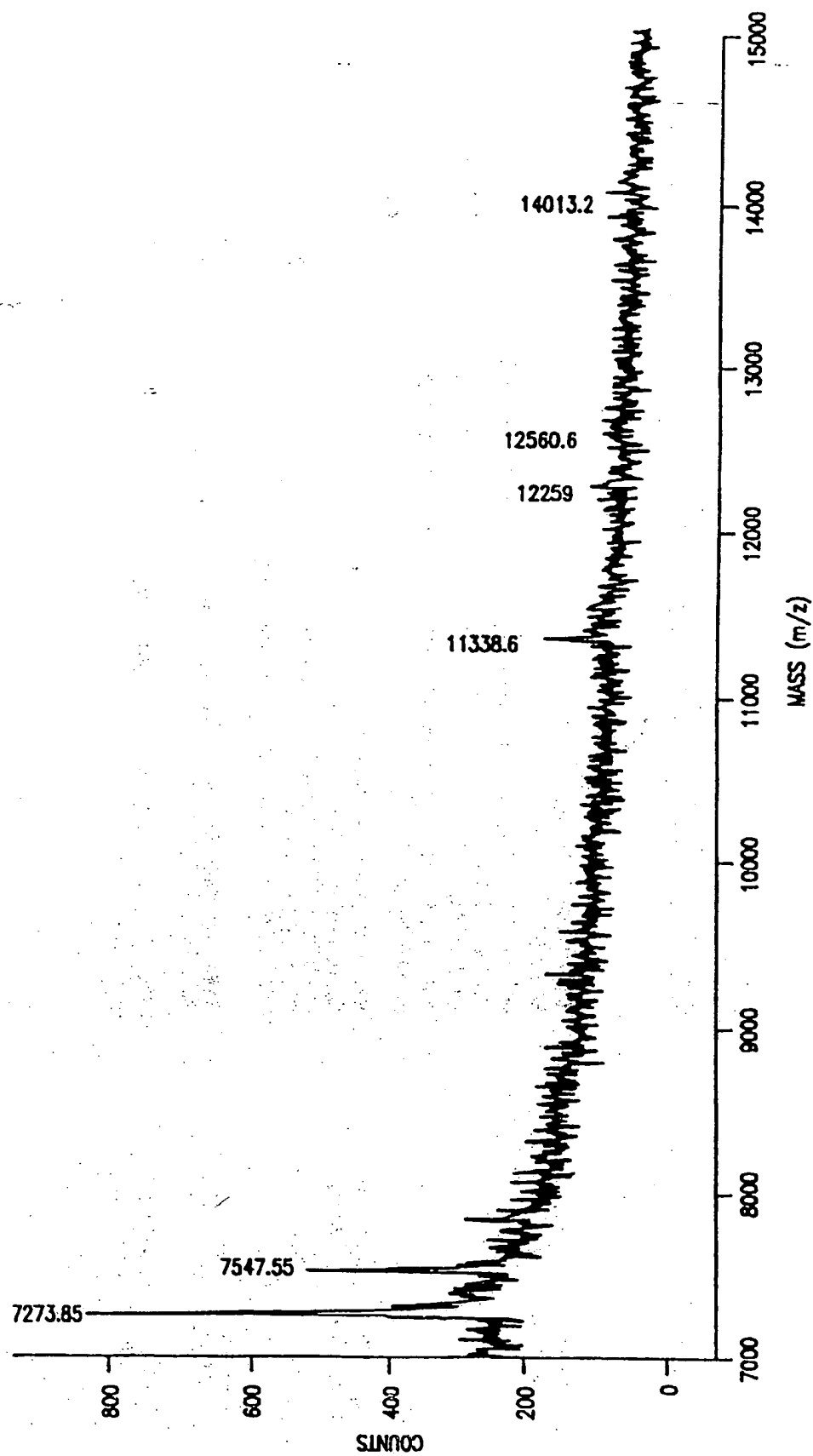


FIG. 86

107/123

## PCR PRODUCT

CKR<sub>a</sub>-F

CAGCTCTCAT TTCCATACA GTCAGTATCA ATTCTGGAAG AATTCCAGACA TTAAGAT  
 AGTCATCTTG GGGCT GTCAGAGTA AAGGTATGT CAGTCATAGT TAAGACCTTC TTAAGGTCT CKR<sub>a</sub>-R-BIO  
 GTAATTCTA TCAGTAAAC CCCGA-BIOTIN

FIG. 87A

## SIZE ANALYSES: WILDTYPE

SENSE STRAND W/O A:

CAGCTCTCAT TTCCATACA GTCAGTATCA ATTCTGGAAG AATTCCAGACA CATTAAAGAT  
 AGTCATCTTG GGGCT

75 bp 23038 Da

SENSE STRAND WITH A:

CAGCTCTCAT TTCCATACA GTCAGTATCA ATTCTGGAAG AATTCCAGACA CATTAAAGAT  
 AGTCATCTTG GGGCTA

76 bp 23349 Da

## SIZE ANALYSES: 32 bp DELETION

SENSE STRAND W/O A:

CAGCTCTCAT TTCCATACA TTAAGAT AGTCATCTTG GGGCT

43 bp 13143 Da

SENSE STRAND WITH A:

CAGCTCTCAT TTCCATACA TTAAGAT AGTCATCTTG GGGCTA

44 bp 13456 Da

PROBE ANALYSES: WILDTYPE (ddTTP TERMINATION):

CAGCTCTCAT TTCCATACA GT

22 bp 6604 Da

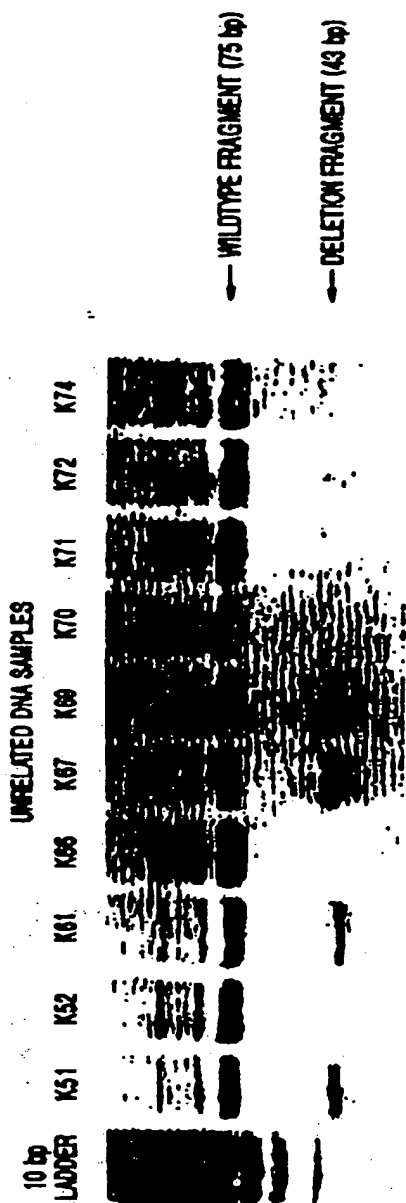
PROBE ANALYSES: 32 bp DELETION (ddTTP TERMINATION):

CAGCTCTCAT TTCCATACA T

21 bp 6275 Da

108/123

DETECTION OF A 32 bp DELETION IN THE CHEMOKINE RECEPTOR BY PCR AND 15% NATIVE PAGE



DELETION FRAGMENTS ARE ONLY WEEKLY STAINED BY SILVER

FIG. 88

109/123

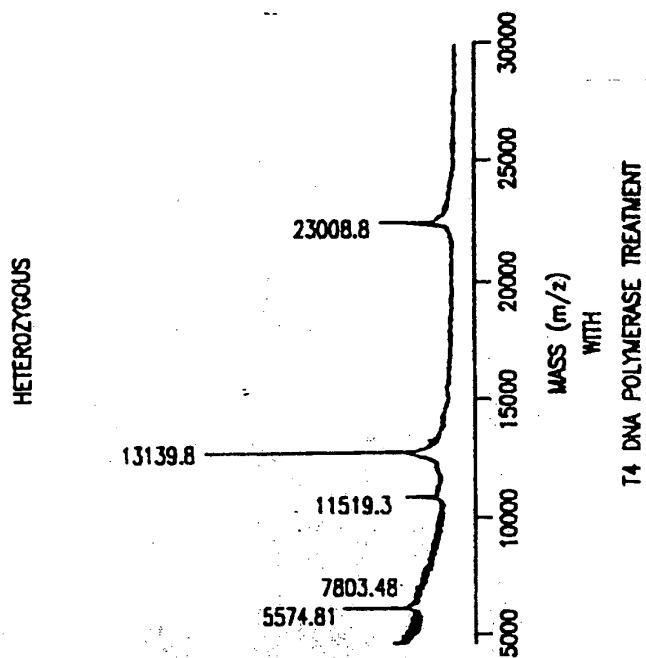


FIG. 89B

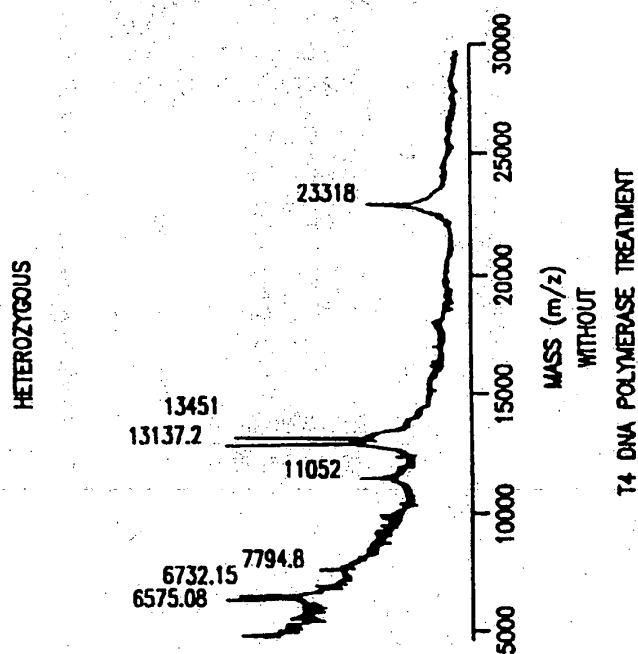


FIG. 89A

110/123

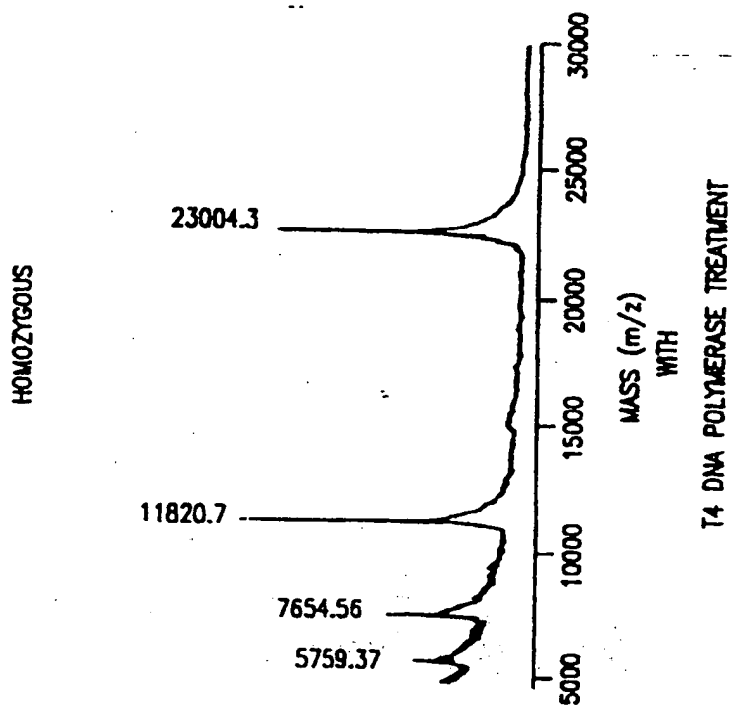


FIG. 89D

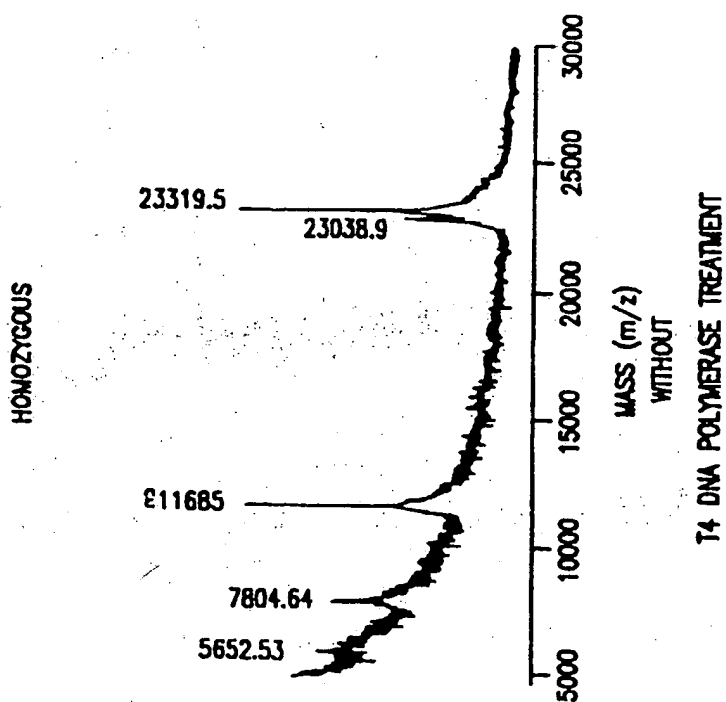
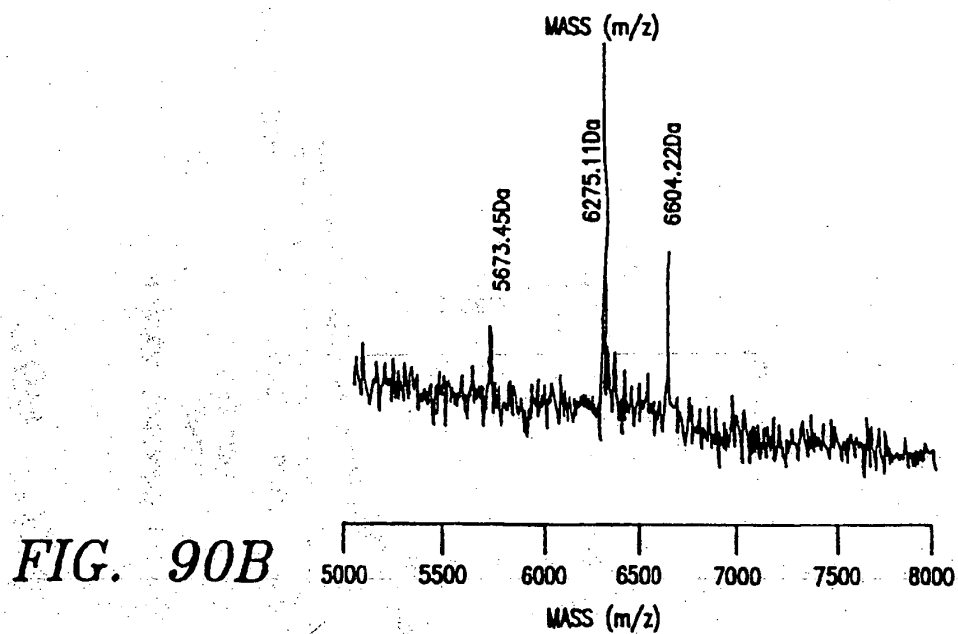
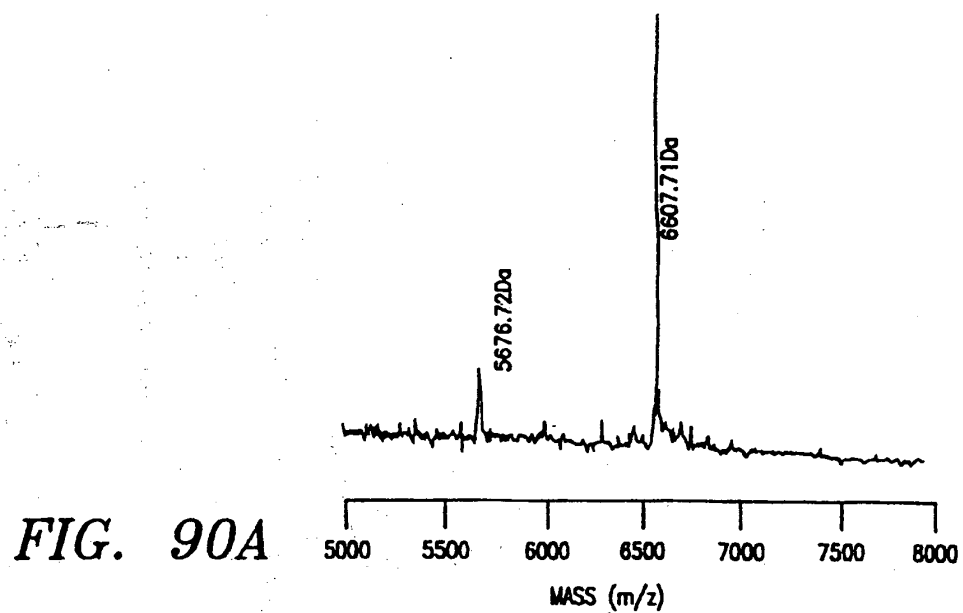


FIG. 89C

111/123



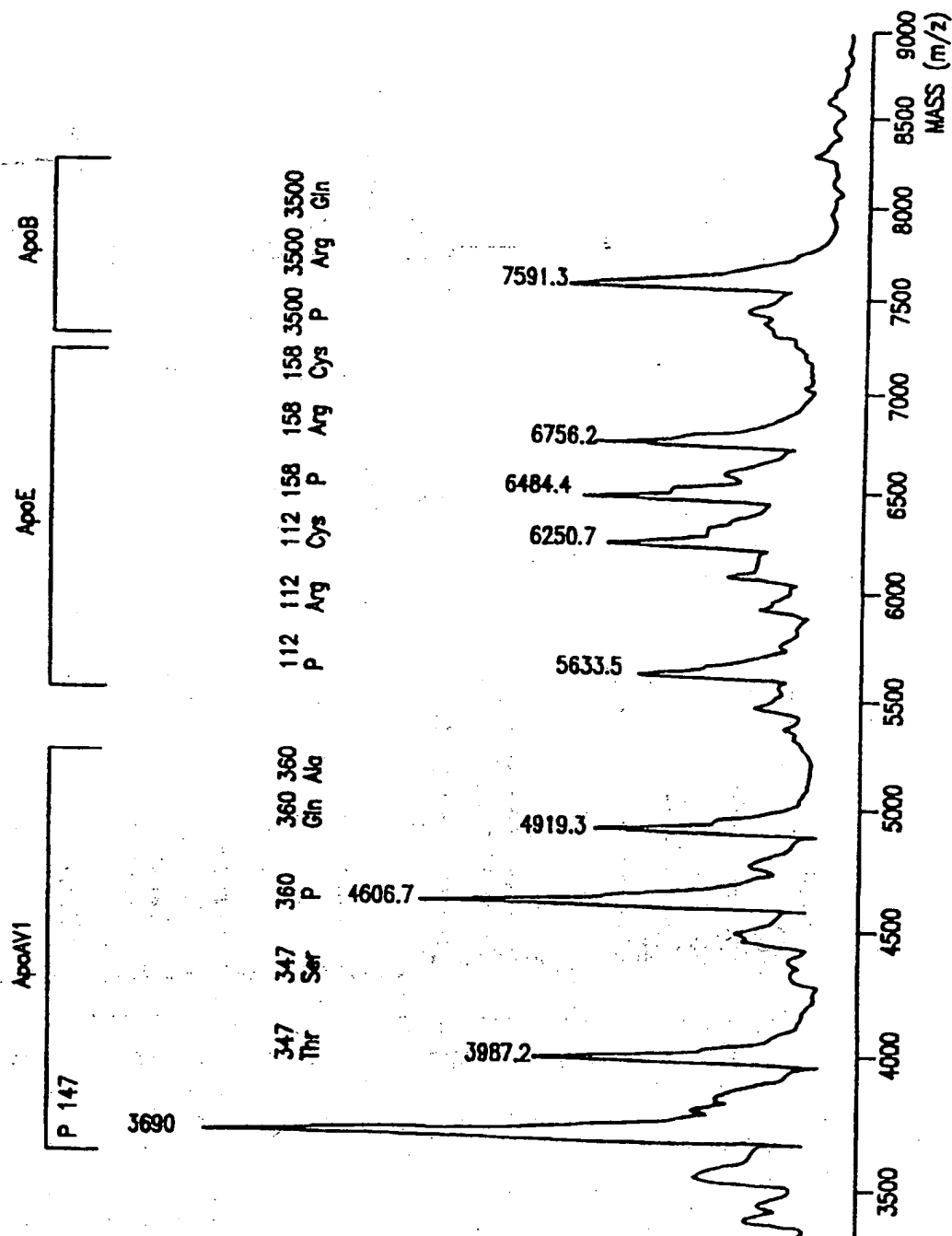


FIG. 91

113/123

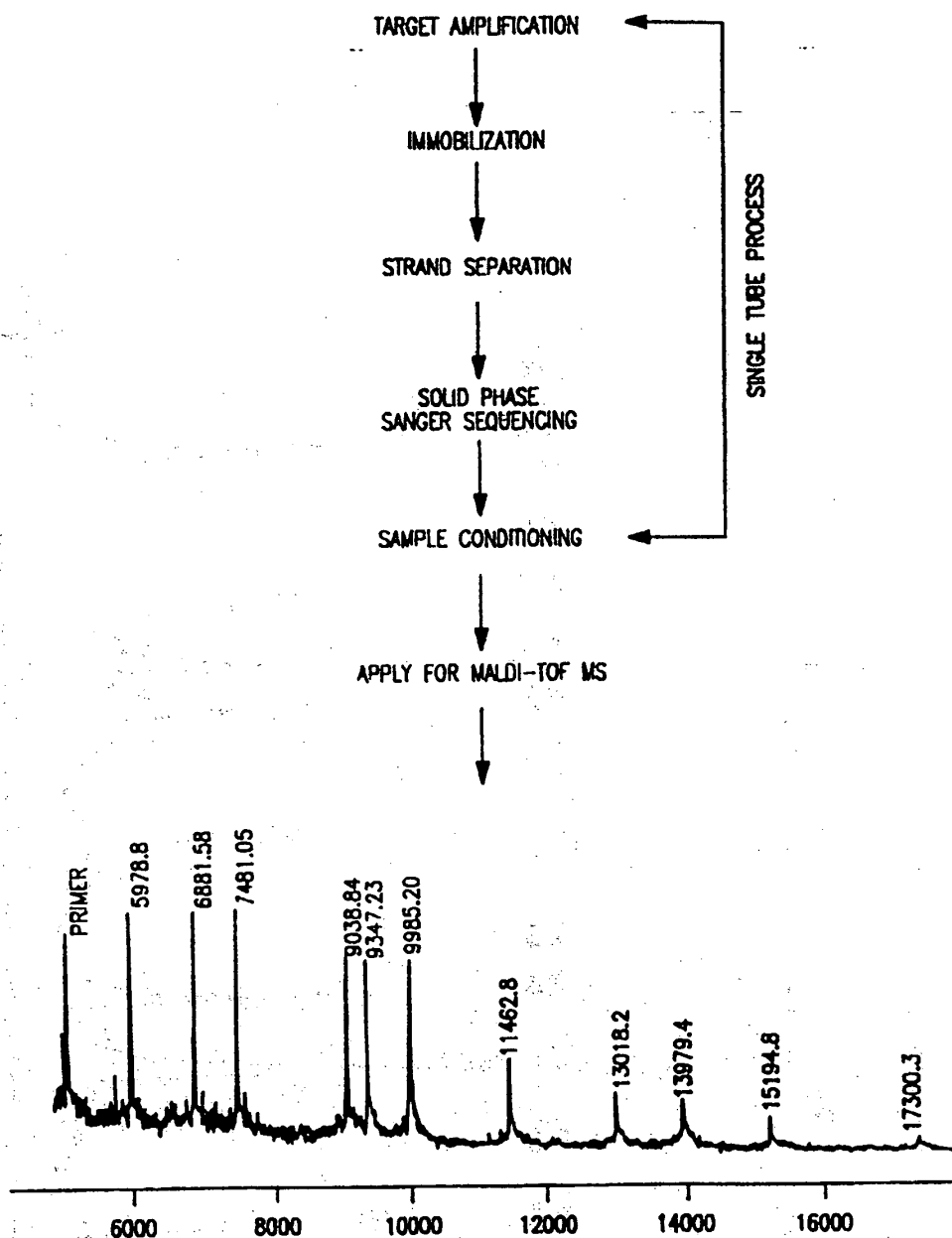


FIG. 92



114/123

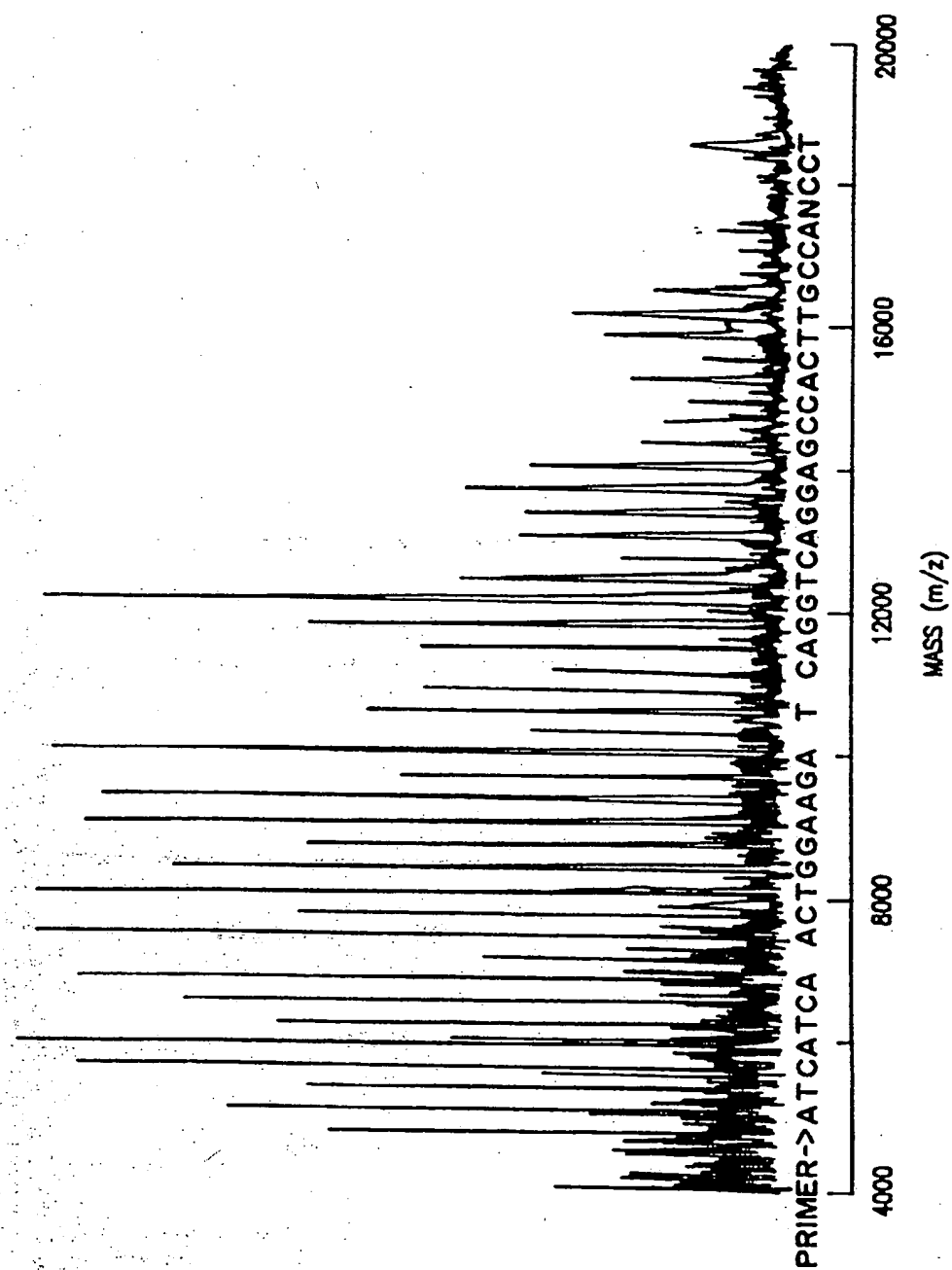
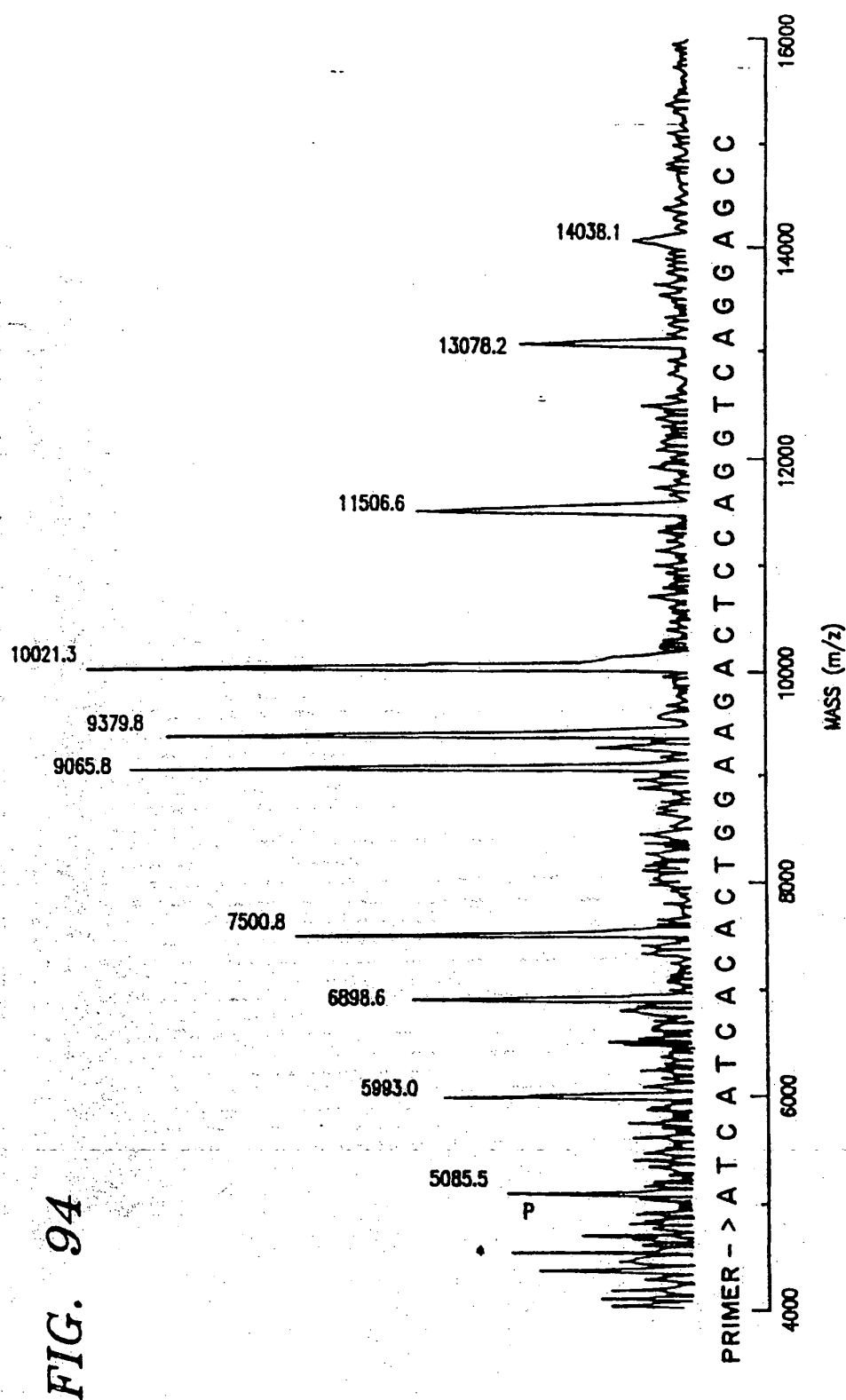


FIG. 93

115/123



116/123

1. MICROTITER PLATE ISOTHERMAL SEQUENCING  
 2. 'PIEZOELECTRIC PIPETTE' TRANSFER  
 3. CORETECH MS  
 P-ATCCACTACACTACATGTGTACAGTTGGwGCwwGddCddC  
 5 10 15 20 25 30 35 40

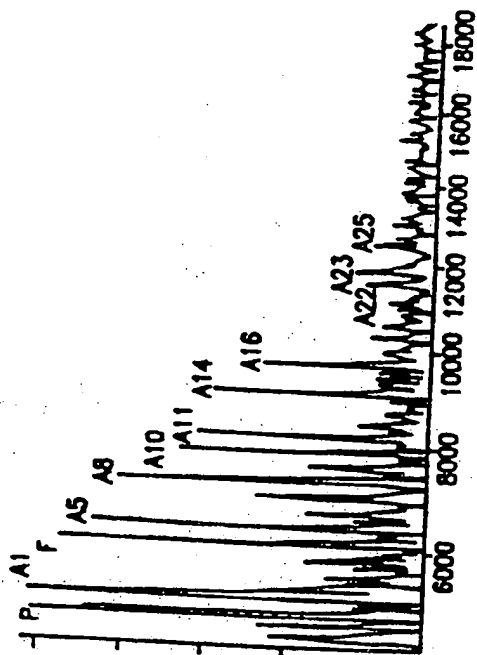


FIG. 95A

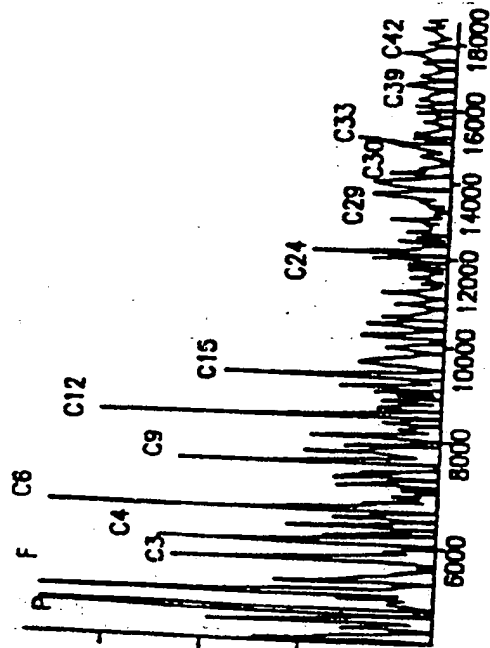


FIG. 95B

117/123

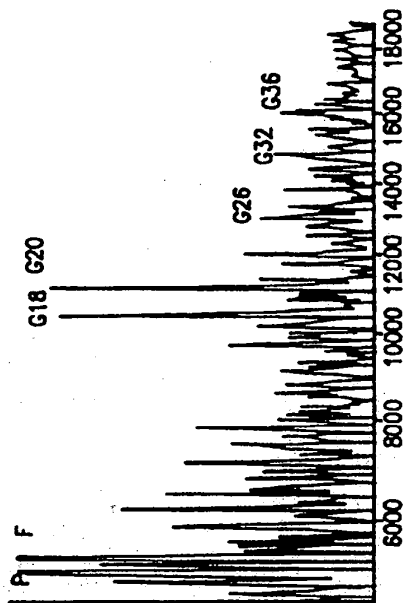


FIG. 95D

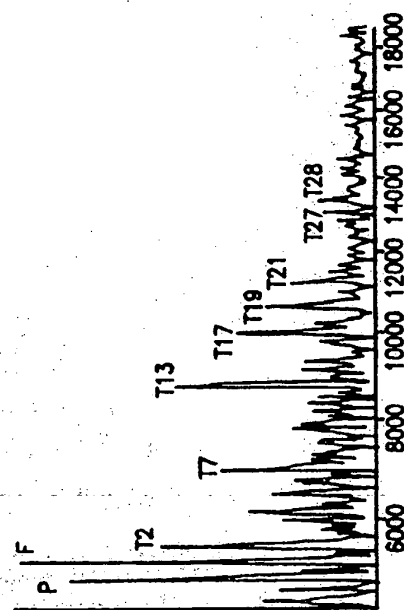


FIG. 95C

118 / 123

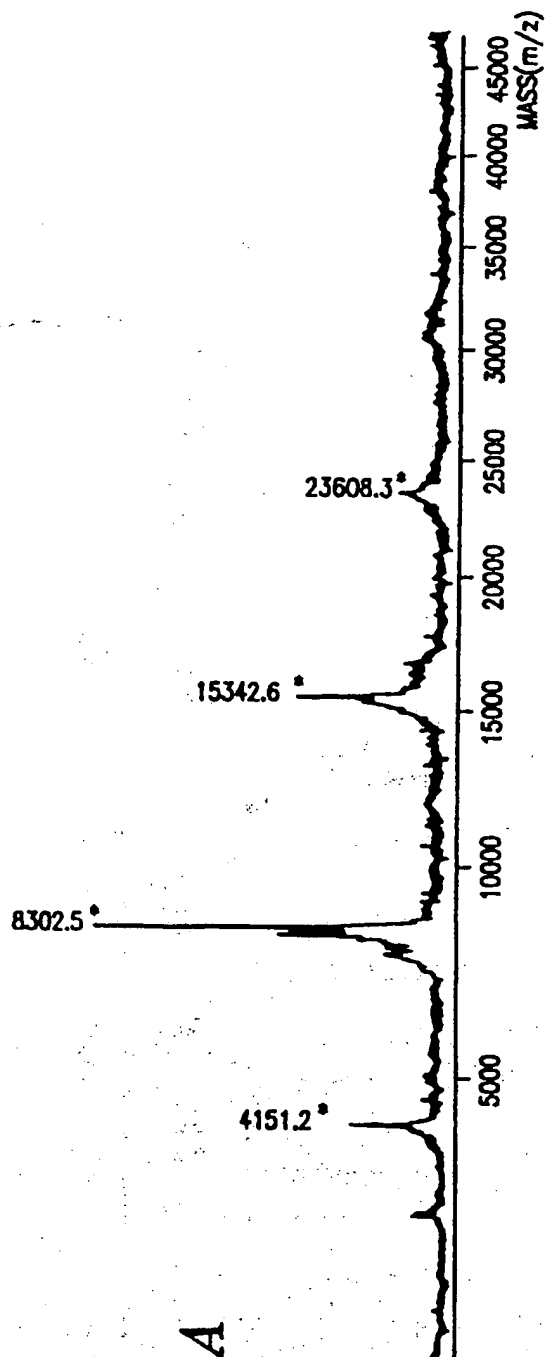


FIG. 96A

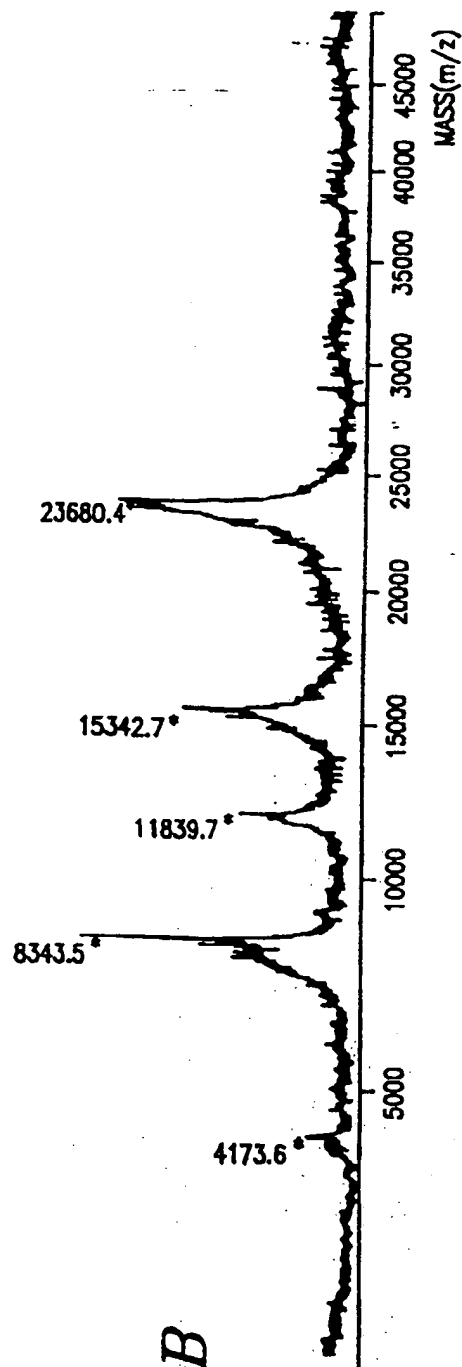


FIG. 96B

119/123

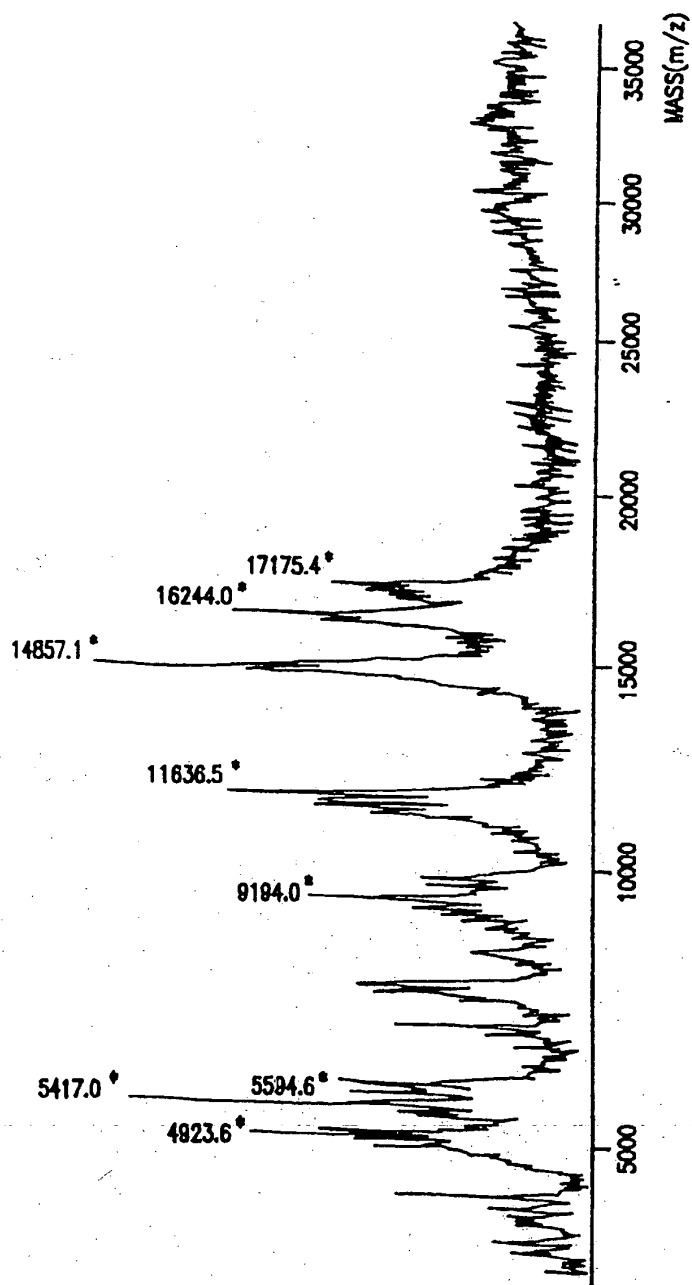


FIG. 97A

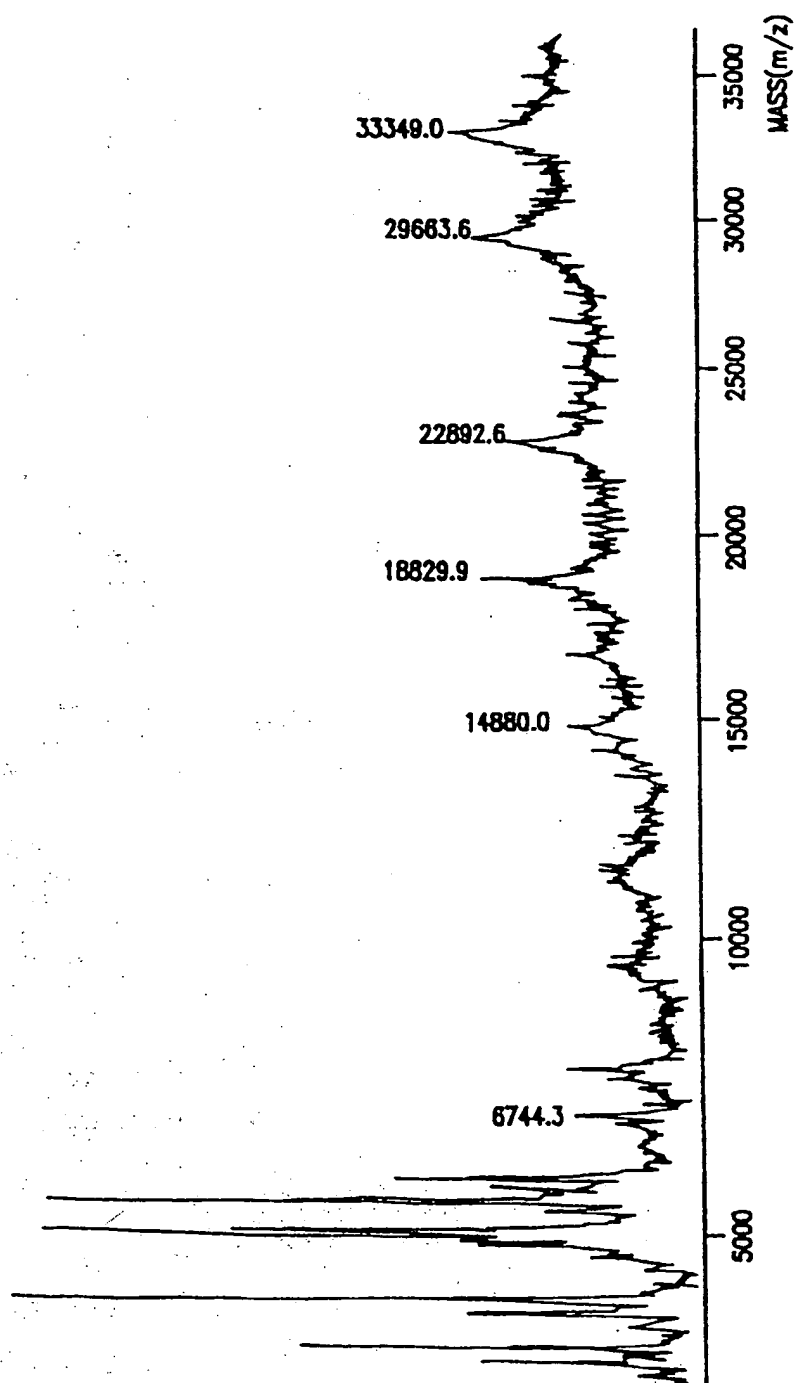


FIG. 97B

121/123

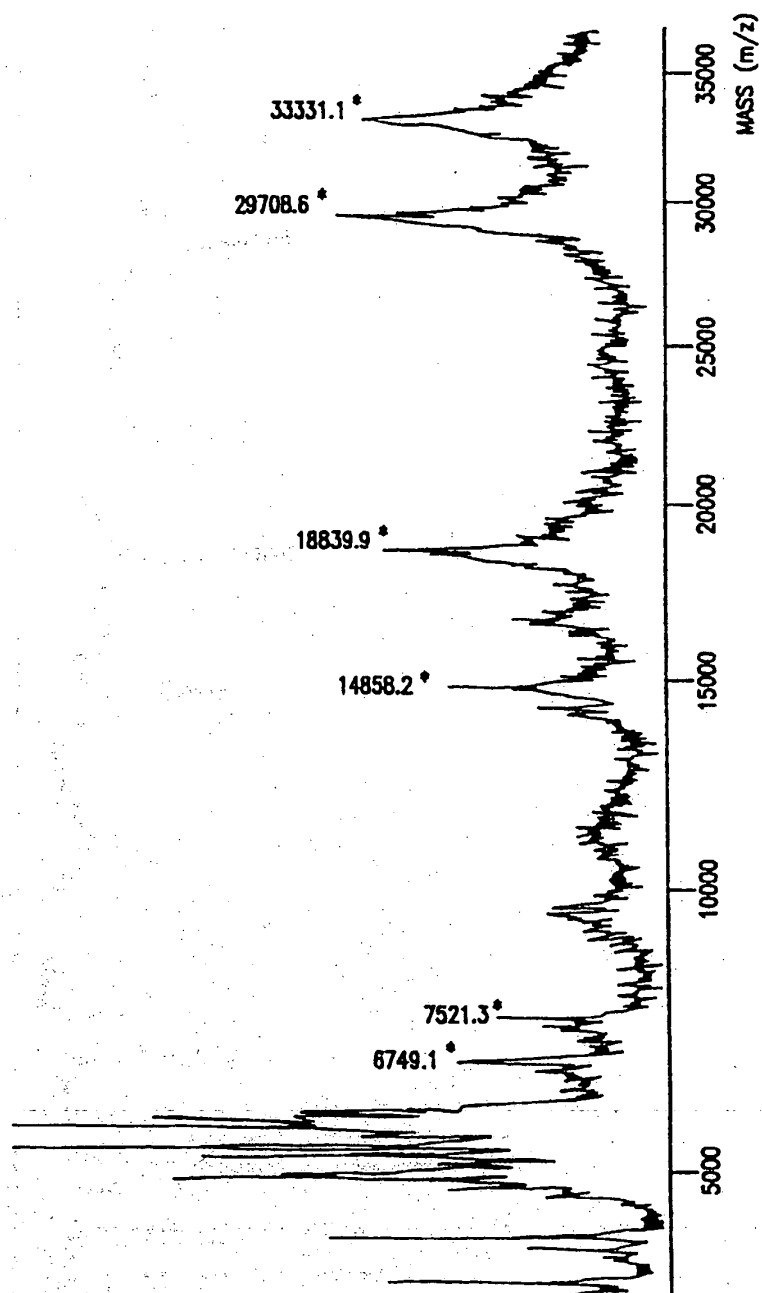


FIG. 98



122/123

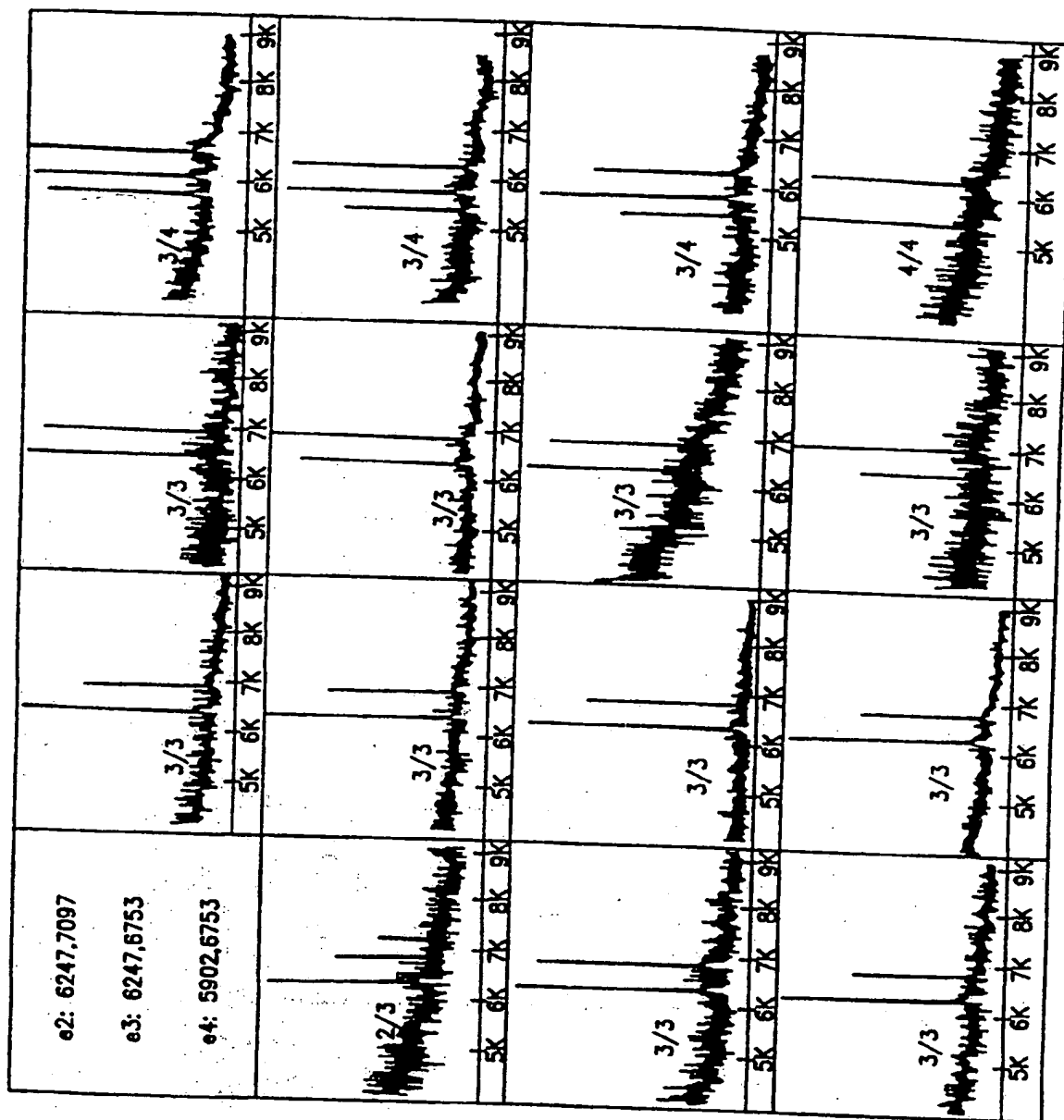


FIG. 99

Apo E  
 CODONS 112&158  
 • ISOTHERMAL PROBE  
 • PTM TRANSFER (~10nL)  
 • MALDI-MS

70% EXPECTED  
 ALLELIC FREQUENCY

123/123

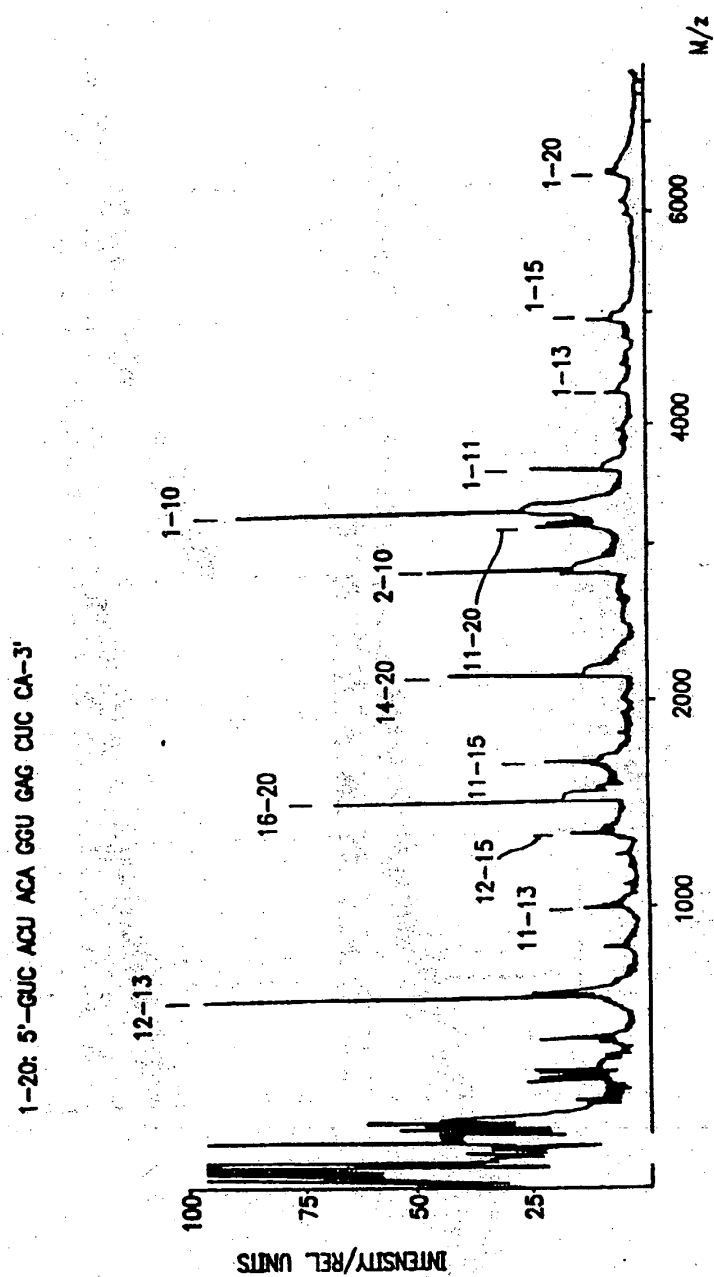


FIG. 100